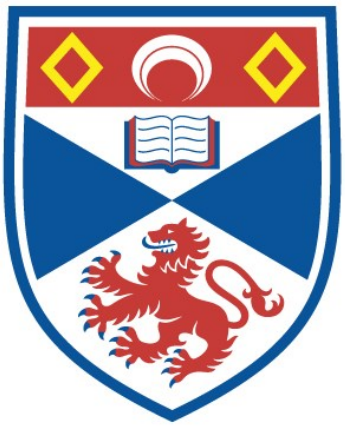
PROTEIN-CARBOHYDRATE RECOGNITION

Stephen Andrew McMahon

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1999

PROTEIN – CARBOHYDRATE RECOGNITION

Stephen Andrew McMahon

A thesis submitted for the degree of

Doctor of Philosophy

20th July 1999

Contents

List of Figures.

List of Tables

List of Abbreviations

AIDS	Acquired immune deficiency syndrome
Ara	Arabinose
CCD	Charged coupled device
CNS	Crystallography and NMR systems
DNase	Deoxyribonuclease
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EXAFS	Extended x-ray absorption fine structure
Gal	Galactose
GalE	Galactose epimerase
Glc-1-P	Glucose 1 phosphate
HEPES	(N- [2-Hydroxyethyl] piperazine-N'-[2-ethanesulphonic acid])
HgOAc	Mercury acetate
IEF-PAGE	Isoelectric focusing polyacrylamide gel electrophoresis
IPTG	Isopropyl β-D- thiogalactopyranoside
K_a	Association constant
Km	Michaelis constant
KPi	Potassium phosphate
Lac NAc	N-acetyl lactosamine
MES	2-[N-Morpholino]ethanesulphonic acid

MPD	2-methyl-2, 4-pentanediol
Myc	Mycolic
NAD(H)	Nicotinamide adenine dinucleotide
NAD(P)	Nicotinamide adenine dinucleotide (phosphate)
NaOAc	Sodium acetate
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
P	Phosphate
PL	Phosopholipid
PDB	Protein data bank
PEG	Poly ethylene glycol
PEG MME	Poly ethylene glycol mono-methyl ether
PMSF	Phenylmethylsulphonyl fluoride
Rha	Rhamose
Rib	Ribulose
RNA	Ribonucleic acid
RNase A	Ribonuclease A
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TDP	Tymidine di phosphate
TRIS	Tris(hydroxymethyl)aminomethane
UDP	Uridine diphosphate
UDP-GlcNAc	Uridine diphosphate N-acetyl glucosamine
UrOAc	Uranium acetate

Vmax	Maximal velocity
WbbL	Rhamnosyltransferase

Declaration.

I, Stephen Andrew McMahon, hereby certify that this thesis, which is approximately 42,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date 20.8.99 Signature of candidate

I was admitted as a research student in September 1995 and as a candidate for the degree of Ph.D. in September 1996; the higher study for which this is a record was carried out in the University of St. Andrews between 1995 and 1998.

Date 20.8.99 Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 20.8.99 Signature of supervisor

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date 20-8-99 Signature of candidate

Abstract

Protein – carbohydrate recognition is an important target for inhibitor development. Improved inhibitor design requires a fundamental molecular basis of these interactions. This thesis describes the preliminary structural studies on three carbohydrate processing enzymes, UDP-galactopyranose mutase, α-D-glucose-1-phosphate thymidylyltransferase and TDP-glucose 4,6-dehydratase. These enzymes are found in important human pathogens such as *Mycobacterium tuberculosis* and *Salmonella typhimurium.*

The major focus of the thesis has been on UDP-galactopyranose mutase, the enzyme responsible for catalysing synthesis of the thermodynamically unfavourable 5 membered ring form of galactose, UDP-galactofuranose from the thermodynamically favoured 6 membered ring form, UDP-galactopyranose. UDP-galactofuranose plays a key role in mycobacterial cell walls.

This thesis also describes work with concanavalin A. This legume lectin is an invaluable model for the study of protein – carbohydrate interactions. Two concanavalin A complexes are discussed. Both structures clear up misunderstandings in the literature and provide an insight into designing enzyme inhibitors.

Acknowledgements

As is tradition in these matters the first person I would like to show my appreciation of is my supervisor, Jim Naismith. His patience, encouragement and enthusiasm are unparalleled and together with his guidance I have arrived at where I am today. For this I am grateful and I thank him.

Throughout my PhD I have called upon many more experienced than myself. For listening and explaining I thank Rob Field, Charlie Weller and particularly Gordon Leonard at the E.S.R.F. whose assistance and time I am especially grateful.

In addition, I would also like to express my gratitude towards Bill Hunter for making his R-axis available, James Matthews for teaching me the fundamentals of molecular biology and Ron Hay for allowing me access to his lab for this purpose. I am also indebted to Phil Evans for trusting me with his xenon equipment and Andy Freer for introducing me to protein crystallography in the first instance.

Sincere thanks go to everyone I have worked alongside throughout this PhD. From post-docs to students, I thank you for making the lab an invigorating environment in which to work. I often relied upon your expertise and sought your input. With reference to this I thank Marie-France Giraud for assistance over the last two years and David Sanders for more recent advice. I would also like to individually thank Huanting Liu and Simon Allard for keeping me entertained, along with the undergraduate students Adam, Charlotte and big

Steve whose assistance I am also grateful. Finally, Davina is due special thanks here. Her companionship in the lab over the last three years was much appreciated.

I have lost count of the times I have called upon the technical staff to perform above and beyond what should be expected of them. For their help when I called upon them I thank Bobby Cathcart, Jim Rennie, Brian Walker and Marge Parker.

When not in the lab I was more than likely playing football for the St. Andrews phenomenon who are Steaua Needarest. For making this such a pleasurable and successful pastime I am in debt to Paul (bostik boot) Baguley, John Devine, Mike Scott, Mike Tathum, Giles Newbury, Andy (The Cat) McCarroll and Rich Pagano.

In the course of writing this thesis I have acquired many new friends without whom the thesis would undoubtedly have been completed somewhat quicker. For entertaining me both during and out with working hours I thank the folks downstairs, Brett, Mike, Dave and Joana, Helen from upstairs, Mike from the sports centre, along with the ever present Simon and Giles. A special mention here goes to Lesley, just thanks!

Penultimateley, I thank all inhabitants and visitors of "12 Balrymonth" and "7 Leonard", particularly John (Werz the Burdz) Devine for his privileged friendship. We had a laugh!

Finally, I am indebted to my parents and family. Without the unconditional love and support I have received from them I would not be where I am today. I only hope I can repay them one day.

For my grandparents……

Chapter 1

Tuberculosis and the Mycobacterial Cell Wall.

1.1 Summary

Tuberculosis is a dangerous disease caused by infection with *Mycobacterium tuberculosis*. It is responsible for around 3 million deaths annually and is estimated to infect as many as one third of the world's population. Established methods of treating infection are in increasing danger of becoming obsolete as new multi-drug resistant strains of the disease emerge. These factors combine to make the requirement for new anti-tuberculosis drugs urgent.

One method of destroying the bacterium is to disrupt the cell wall. This can be achieved by targeting the enzymes responsible for cell wall biosynthesis. The mycobacterial cell wall is uniquely complex: its biosynthesis affords novel targets for drug intervention.

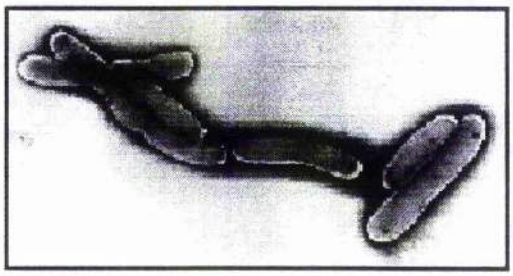
The successful development of new drugs is greatly aided by a three-dimensional picture of the target enzyme active site. This can be obtained by protein crystallography. This chapter describes the re-emergence of the mycobacterial infection tuberculosis and the strategies being adopted to over come this. Particular emphasis is placed on the structure and biosynthesis of the mycobacterial cell wall and the targets it presents for novel drug intervention.

1.2 Introduction

There are over 70 different strains of mycobacteria. All are non-capsulated, acid fast, obligate aerobes with characteristic cell walls containing an unusually high percentage of lipids. Most appear as harmless saprophytic inhabitants of soil, however there are notable exceptions, the most important of which is *Mycobacterium tuberculosis* (*M. tuberculosis*), Figure 1.1. This infects one third of the worlds population (Dolin *et al*, 1998). Other pathogenic mycobacteria include *Mycobacterium leprae* (*M. leprae*) which causes leprosy, and a class of rarely occurring species termed the 'Atypical Mycobacteria'. The remaining mycobacteria are primarily environmental organisms collectively known as MOTTS (Mycobacteria Other Than Tuberculosis) (Elliot *et al*, 1997).

Figure 1.1: M. tuberculosis.

(http://www.act.uct.ac.za/depts/mmi/lsteyn/lecture.html)



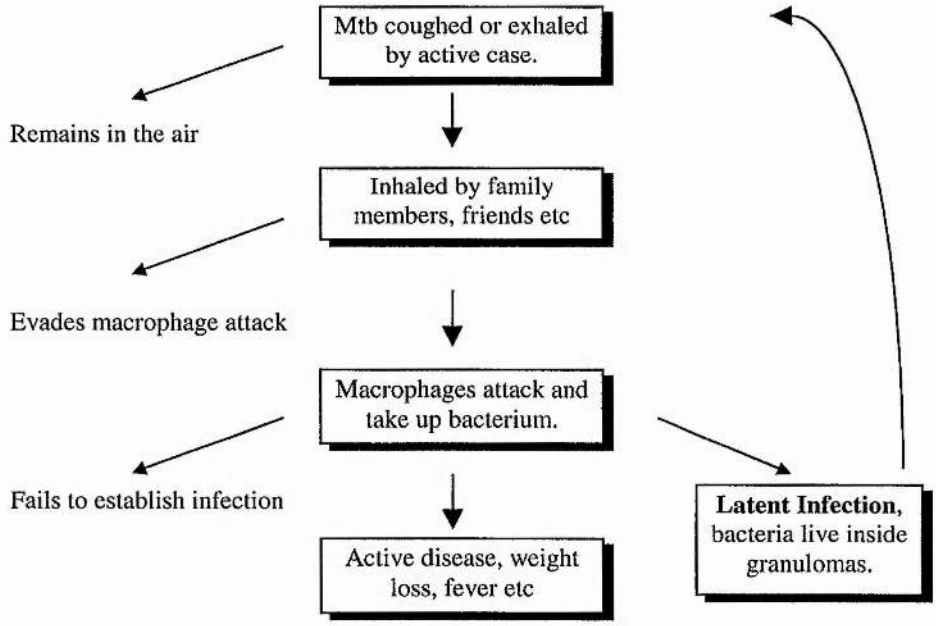
Pathogenesis of Tuberculosis

Pulmonary tuberculosis (TB) is the most common form of the disease and is transmitted in the form of aerosol droplets that may be inhaled by the lungs. The most obvious symptoms of pulmonary TB include fatigue, weight loss, chest pains and night fevers. Once the bacterium has gained entry to the body it activates the body's defence mechanisms and is

engulfed by macrophages. Macrophages phagocytose foreign objects and destroy them upon release of toxic enzymes. For reasons not yet fully understood *M. tuberculosis* infections evade this attack and survive. If the invading bacterium survives it can follow one of two available routes proceeding to latent infection or active disease. In both scenarios the macrophages fail to eradicate the infection and it is the immune system's ability to contain the bacterium which determines the route the infection. In the clinical disease, the mycobacteria divide inside the macrophages until the macrophage ruptures spilling not only the mycobacteria but also toxic enzymes. This process results in lesions in the lung where the mycobacteria multiply freely, eventually destroying the lung tissue and rupturing blood vessels; this is the active form of the disease and results in the blood stained sputum.

Alternatively the host immune system may halt the spread of the disease by containing infected macrophages behind a barrier of white blood cells and phagocytes known as granulomas. This is the latent form of the disease, and unless the immune system is compromised represents no danger to the individual. However, if the immune system becomes suppressed for example as result of an accident, viral infection etc. the otherwise latent mycobacterial infection may proceed to the clinical form of the disease, Figure 1.2.

Figure 1.2. The pathogenesis of tuberculosis



History of Tuberculosis

M. tuberculosis was first identified as the bacterium responsible for tuberculosis in 1882 (Koch, 1882), although it is thought the disease has been around for many centuries prior to this discovery. Evidence for these claims was uncovered in 1994 when Salo and co-workers recovered DNA unique to *M. tuberculosis* from a lung lesion of a spontaneously mummified 1000 year old adult female body in southern Peru (Salo *et al*, 1994).

The incidence of the disease in between these centuries is unclear. It is believed that almost all Europeans were infected with TB in the 17th century, with the bacterium being responsible for one in four deaths at this time (http://www.brown.edu/Research/TB-HIV_Lab). Over the next century it is thought the incidence of tuberculosis went into decline before a resurgence to peak levels in the early 1800's, when the annual mortality rate in Europe and the U. S. was around 5000 per 1 000 000. The number of tuberculosis infections then went into decline, and in 1882, when Koch gave his historic lecture, its

occurrence had fallen over 50% from the peak levels reported in the early 1800's. This trend continued into the 20th century and was accelerated mid-century with the introduction of effective therapies to combat the disease. In 1969 the TB death rate in the U. S. was down to 40 per 1 000 000 (http://www.brown.edu/Research/TB-HIV_Lab). Perhaps inevitably the last two decades have seen this trend reversed and incidences of TB have again started to rise throughout the whole world.

Current Status of TB

In 1997 more people died from TB that any year in living history (World Health Organisation, 1998). A disease once thought to be confined to the realms of history, TB is most definitely back with a vengeance. That year there were approximately 7.25million new infections of TB and around 3 million deaths as a direct result of TB infection (World Health Organisation, 1998). This is more than the combined total from malaria, diarrhoea, HIV/AIDS and tropical diseases, conspiring to make TB one of today's most pressing global problems. In addition one in every three people are hosts to a dormant TB infection, (each has a 10% chance of progressing from infection to clinical disease), providing a significant reservoir of disease for the future. Moreover, the disease is a world-wide problem and is not confined to poor underdeveloped nations, Table 1.1. TB occurs in every country in the world with the highest incidences being in Asia and Africa. In recent years it has re-emerged as a major heath issue in Russia where as many as 10% of Russian prisoners are infected. More alarmingly, 20 000 of these prisoners are infected with new emerging strains of the disease known as Multi-Drug Resistant - TB or MDR - TB (TB Treatment Observer, 1998). These deadly versions of the disease have developed a resistance to at least two of the

frontline drugs used to treat TB. Moreover in a few isolated cases they have been found to be resistant to as many as seven anti - TB drugs (Frieden, 1996).

Table 1.1: Global distribution of TB cases in 1997 (Dolin et al, 1998)

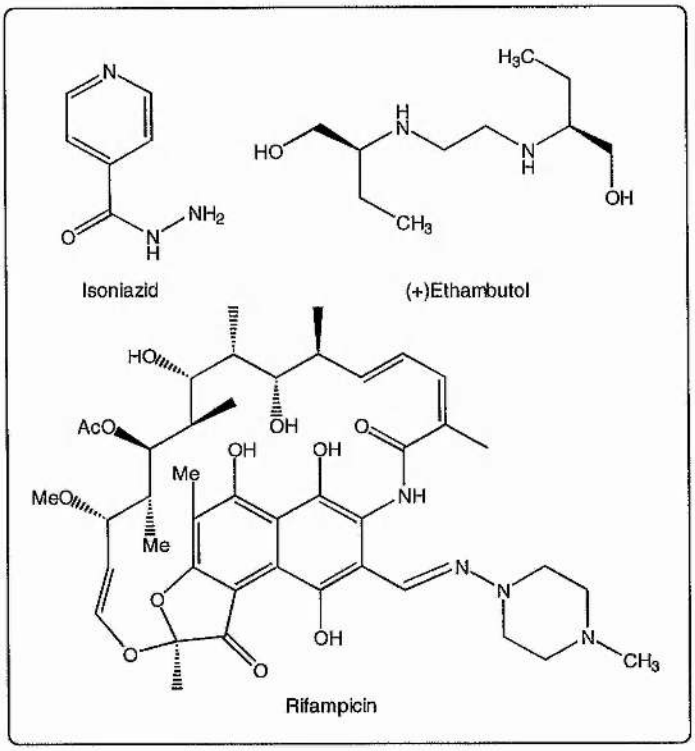
Region	No. of Cases	Deaths
Southwest Asia	2 800 000	1 095 000
Africa	1 650 000	770 000
Western Pacific	1 583 000	594 000
Americas	448 000	160 000
Eastern Mediterranean	427 000	173 000
Europe	342 000	118 000
Total	**7 249 000**	**2 908 000**

Treatment of TB

What is particularly distressing about the current death rates as a result of *M. tuberculosis* infection is that they are largely avoidable. There are several methods of overcoming infection by *M. tuberculosis*, though the efficiency of the therapeutic measures used may vary depending on which area of the world they are applied (Duncan, 1997). The current World Health Organisation (WHO) recommendation is a combination of 4 front-line anti-mycobacterial drugs, isoniazid, ethambutol, pyrazinamide and rifampicin, administered over a minimum period of two months followed up by a further four months of just two of these drugs, isoniazid and rifampicin Figure 1.3. This multi-drug therapy is known as DOTS (Directly Observed Treatment Short course) and has cure rates of upto 95% (Dolin *et al*, 1998). Furthermore, treatment by DOTS is inexpensive, a six month course costs only £25, allowing it to be implemented world-wide as an effective treatment of TB. Attempts to treat

TB with a single drug often result in the selection and growth of drug resistant strains of the bacteria. It was this observation that catalysed the introduction of DOTS.

Figure 1.3: Three of the WHO recommended front-line drugs used to treat TB.



In addition to DOTS a vaccination against TB is also available. The bacille Calmette-Guerin (BCG) was first introduced in the 1950's and involves inoculation with a live attenuated strain of the closely related species *Mycobacterium bovis* (Evans, 1998). The efficacy with which the BCG works seems to depend on the region of the world it is applied. Administration in the UK affords greater than 80% protection against infection whereas in Southern India protection rates are almost zero (Duncan, 1997). As to why this is the case no-one exactly understands. At present, the most favoured explanation involves co-infection with other non-pathogenic bacteria that interferes with, or masks, BCG protection (Bloom & Fine, 1994).

Reasons for Resurgence of TB

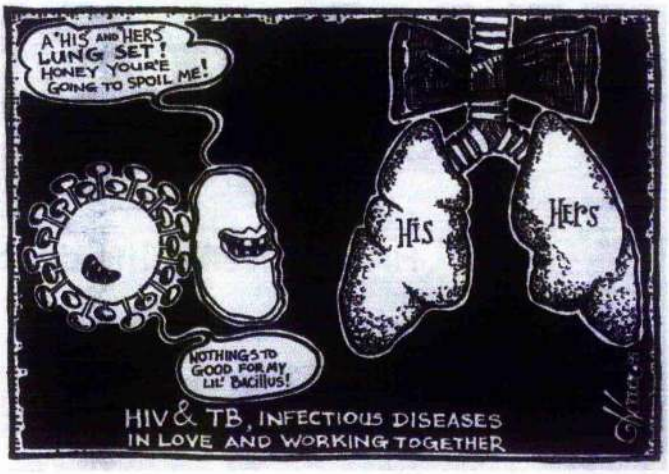
Despite the existence of effective TB therapies, the recent rise in numbers of TB infections is a problem of human making. There are many reasons for the high number of unnecessary deaths including a lack of easily accessible health services, diagnostic facilities, delays in patients seeking medical attention and perhaps most importantly, the failure of patients to comply with prescribed regimens.

The emergence of MDR-TB is primarily a result of insufficient resources and poor patient compliance, that is, failing to complete a full course of therapy. This provides the selection mechanism needed for the development of resistance in mycobacteria. MDR-TB is profoundly more difficult to cure than drug sensitive mycobacteria, requiring more expensive and toxic drugs for treatment. Even then it is often the case that treatment is unsuccessful. It does not seem a coincidence that the regions throughout the world with a high prevalence of MDR-TB are those with a poor health care infra-structure leading to poor patient compliance

(WHO, 1998). Strains of MDR-TB have been known to be resistant to as many as 7 different classes of anti-mycobacterial drugs resulting in mortality rates of patients infected being on a par with the pre-antibiotic era.

The Human Immunodeficiency Virus (HIV) epidemic represents an additional contributory factor to the resurgence of TB. Infections with HIV compromise the host's immune system increasing the likelihood, by about 10% per year, that a dormant mycobacterial infection will start to replicate and progress to the active form of the disease (Young, 1995). In turn, the development of an active mycobacterial infection speeds up the progression of HIV to AIDS. Considering approximately one third of the world's population is host to a dormant mycobacterial infection, the synergy between HIV and TB, Figure 1.4, is likely to result in substantially increased numbers of TB infections. The WHO estimates the HIV/TB link will result in an additional 750 000 cases of TB annually which otherwise would not have occurred (TB Treatment Observer, 1998).

Figure 1.4: Cartoon representation, taken from the WHO's 'TB treatment Observer', of the co-operation displayed between TB and AIDS.



Unfortunately research into *M. tuberculosis,* and the way it causes disease has been a low priority over the past thirty years. An over reliance on the primary frontline drugs, together with the belief that the disease was confined to the third world saw research funds restricted, and little progress made in the research and development of new anti-mycobacterial drugs. Furthermore, as *M. tuberculosis* is an extremely difficult and dangerous pathogen to work with, requiring specialised expensive containment facilities, it was not a particularly appealing organism to be working on. In addition, as *M. tuberculosis* cells divide only once in 24 hours it takes three to four weeks to form a single colony on an agar plate under ideal incubation conditions. This results in significant delays in diagnosis when the infection can be freely transmitted in the interim period.

Against this backdrop, the urgency with which new anti-mycobacterial drugs are required is clear. Accordingly, recent years have seen a substantial increase in the funds being made available to achieve this goal. Several lines of novel research are being pursued, including genomics, immunotherapy, vaccination and the targeting of the mycobacterial cell wall in the quest for new anti-mycobacterial agents.

1.3 New Developments in Tuberculosis Treatment

Immunotherapy

This is an alternative line of approach, in that the host is targeted rather than the pathogen. Patients are treated with a heat-killed preparation of the non-pathogenic *Mycobacterium vaccae* before embarking on a course of DOTS (Stanford *et al*, 1990). Whilst this failed to improve the overall cure rate, significant improvements were made in the treatment of MDR-TB or where patients had failed to complete the prescribed drug course (Etemadi *et al*, 1992). Why this is so is not exactly clear, but it is thought that an immune response is stimulated allowing the host to eliminate latent mycobacteria which are unaffected by the drugs. This method of treatment is currently undergoing phase II clinical trials in South Africa.

Vaccines

Progress is being made in different areas towards generating new candidate vaccines. Research has primarily focused on development of attenuated strains of *M. tuberculosis* (McAdam *et al*, 1995), protein subunit vaccinations (Anderson *et al*, 1995), nucleic acid vaccines (Ulmer *et al*, 1993) and improving the existing BCG vaccination (Maharias *et al*, 1996). However, despite many advances in these areas, any new vaccine would have to be more effective than the existing BCG. For this to be proven a method of monitoring protective immunity must be identified in order to avoid lengthy and expensive clinical trials.

Genomics

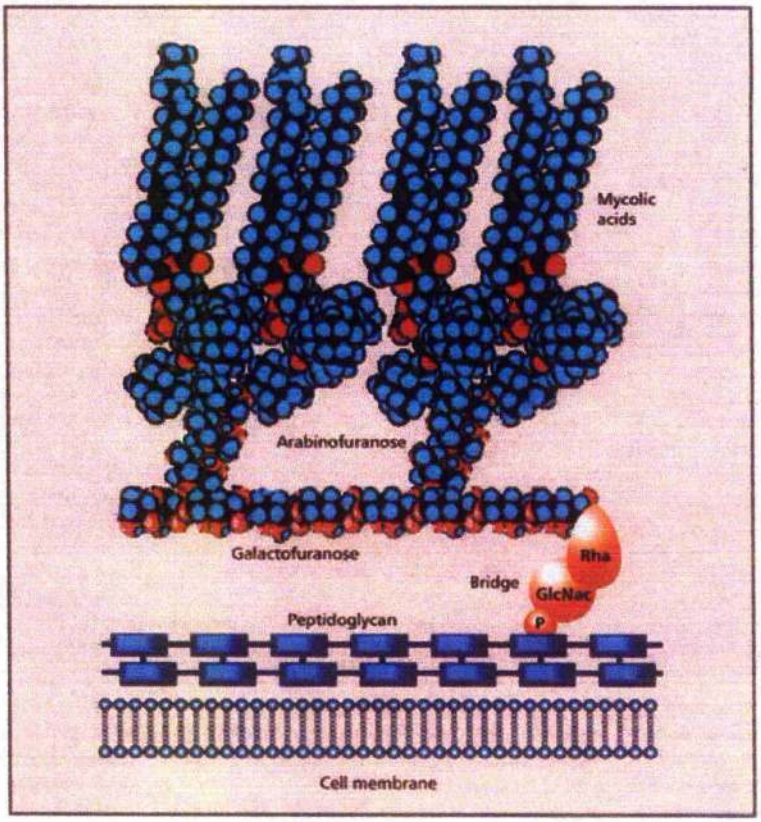
The sequencing of the complete *M. tuberculosis* genome has been completed by Cole *et al* (1998) revealing the sequence of every potential drug target of any new anti-mycobacterial drug. The sequence consists of some 4.4 million base pairs being surpassed only in size by the *E. coli* genome (Blattner *et al*, 1997). The sequence confirms the bacteria to have the metabolic potential to exist in a variety of environments. This may explain how the bacteria survive in what is thought to be a largely anaerobic environment when the infection is housed in the dormant state. Indeed, Cunningham and Spreadbury (1998) have reported a thickening of the mycobacterial cell wall when the bacterium are starved of oxygen, possibly strengthening the bacteria's ability to survive in the dormant state.

1.4 Mycobacterial Cell Wall

At the lecture of Robert Koch in 1882, when *M. tuberculosis* was first identified as the agent responsible for TB, the unusual appearance of the mycobacterial cell wall was commented upon (Koch, 1882). Subsequently with sequencing of the *M. tuberculosis* genome, its importance in mycobacterial physiology has been confirmed with the identification of some 250 genes associated in lipid and polysaccharide metabolism, providing a cornucopia of potential new drug targets.

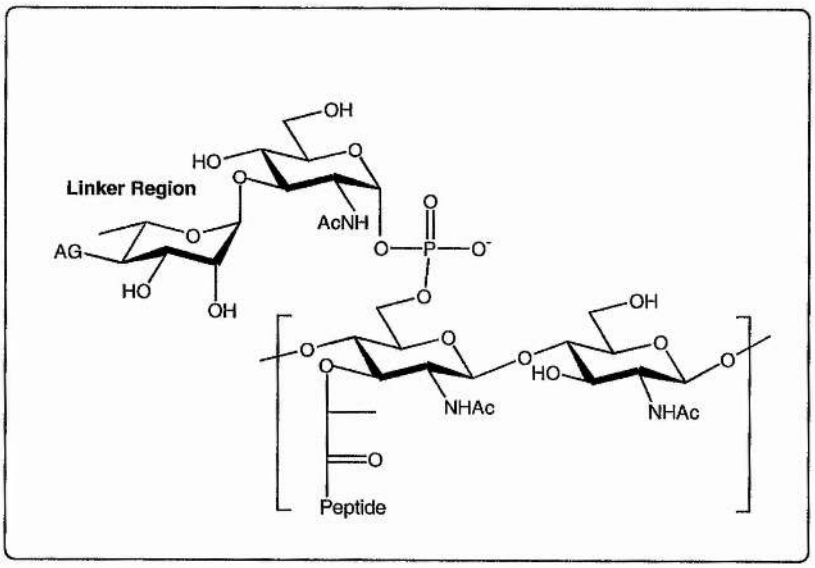
The structure of the cell wall itself is an elaborate one and it presents a formidable barrier to anti - tuberculosis drugs, Figure 1.5. It consists of several unusual polymers and is thick and hydrophilic in nature. This confers on their bacteria an extraordinary resistance to drugs, the host immune system, dehydration and acid treatment (Draper, 1982).

Figure 1.5: The mycobacterial cell envelope. UDP-galactopyranose mutase is located in arabinogalactan and catalyses the interconversion of UDP-gal$_p$ and UDP-gal$_f$. LAM is not shown, (Evans, 1998).



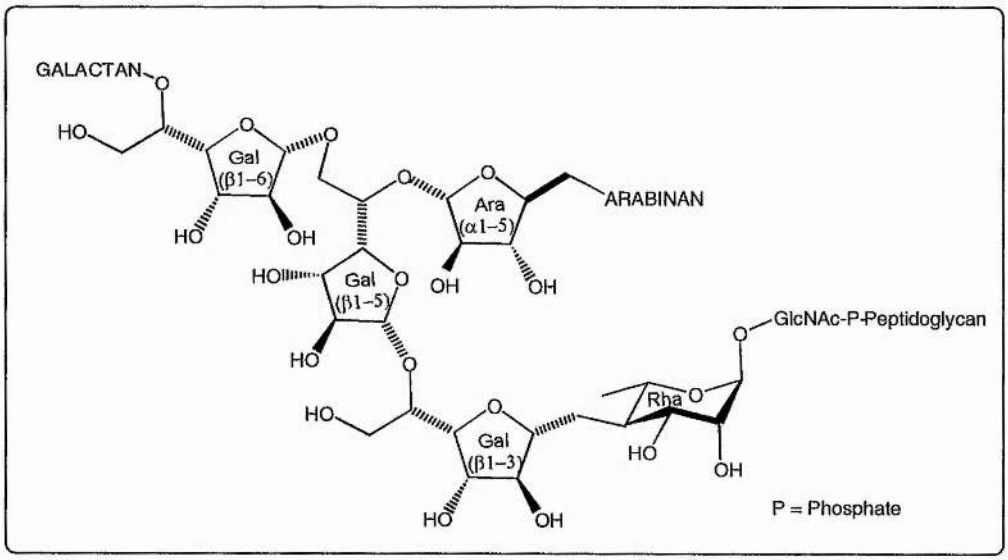
Four classes of polymers make up the cell wall: peptidoglycan, arabinogalactan, mycolic acids and lipoarabinomannan (LAM). Three of these polymers make up what is the fulcrum of the cell wall in the mycolyl arabinogalactan-peptidoglycan complex (mAGP). The first of these four polymers, peptidoglycan, consists of linear polysaccharide chains extensively cross-linked by short peptides. Peptidoglycan is covalently linked to the next polymer in the cell wall, arabinogalactan, via a unique diglycosylphosphoryl bridge containing rhamnose and *N*-acetyl glucosamine, Figure 1.6.

Figure 1.6: The linker region connecting the peptidoglycan and arabinogalactan.



Arabinogalactan is an essential structural component of the cell wall, forming a connection between peptidoglycan and mycolic acids. Its backbone consists of alternating β(1, 5) and β(1, 6) galactofuranose (gal_f) residues branched from which are three units each consisting of linear α(1, 5) linked arabinan. The sugars that make up the arabinogalactan are all in their furanose configuration, (5 membered ring structure), as opposed to the more commonly found 6 membered ring structure (pyranose) Figure 1.7.

Figure 1.7: Simplified representation of the arabinogalactan backbone.

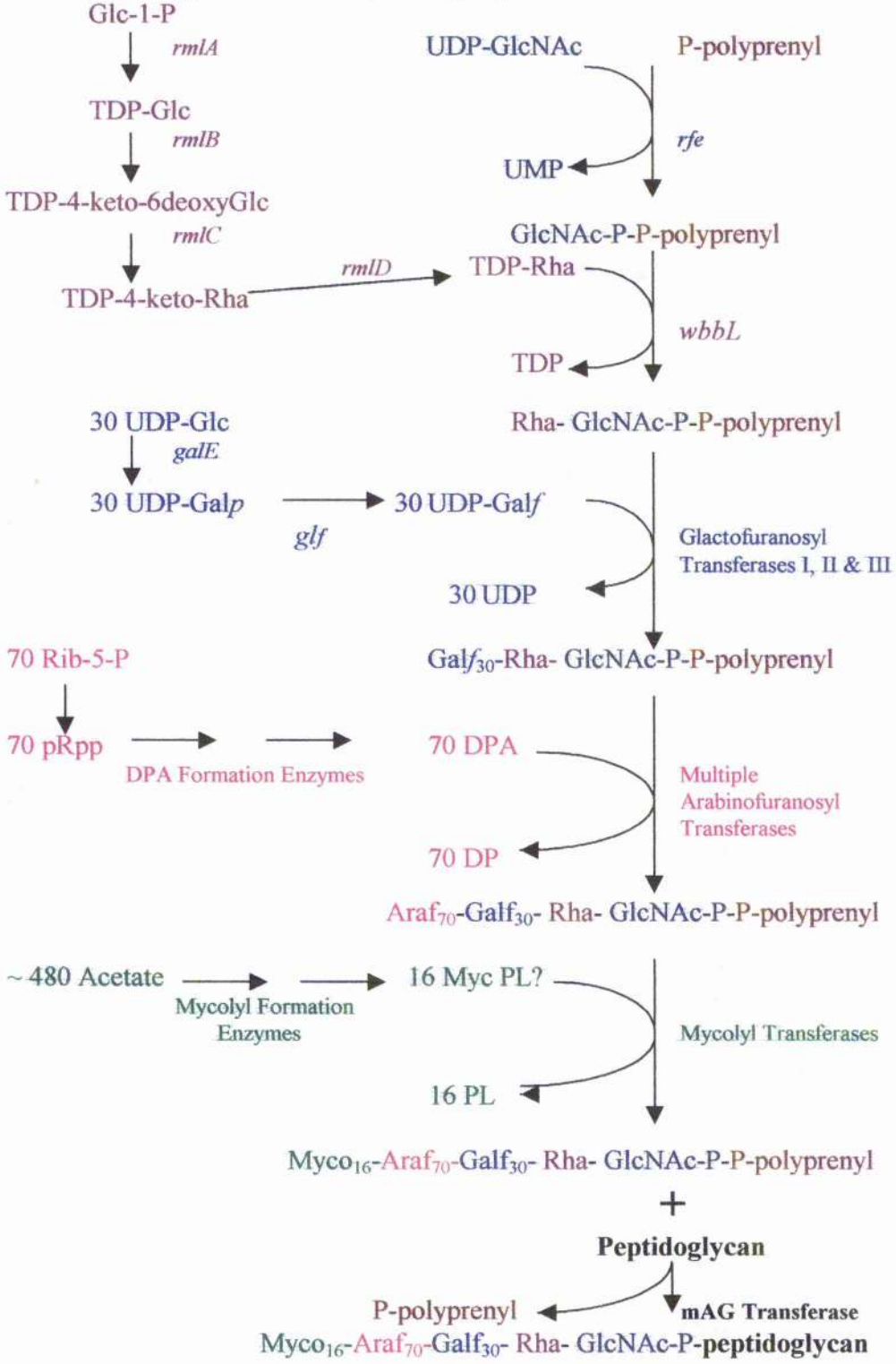


The mycolic acids are located in clusters of 4 on the terminal pentaarabinofurunosyl units, consisting of a mixture of different C_{60} – C_{90} branched fatty acids forming a pseudo – outer membrane, making the membrane a highly hydrophilic structure. Approximately two thirds of the terminal pentaarabinofurunosyl are mycolated. In the mAGP complex, arabinogalactan forms an integral part of the cell wall proper whereas the fourth polymer LAM has been implicated as a key surface molecule involved in the virulence of mycobacterial disease. LAM's complex structure comprises a core of 6 mannose residues, substituted to varying degrees at position 2 with single α-mannose residues, directly attached to position 6 of the inositol of the phosphatidyl linker. To date it is still unclear how LAM is associated with the cell wall; though unpublished results (P. Draper) imply the lipoglycan is firmly but not covalently attached. In addition, monoclonal antibodies to LAM recognise whole mycobacterial cells in enzyme linked immune assays suggesting at least part of the molecule is accessible to the outside environment (Chatterjee, 1997). Structurally, the

macromolecular arabinan is an ensemble of oligomeric α(1, 5)-D-arabinofuranose chains via bridging α-D-arabinofuranose residues at numerous branched points, terminating with a β2-D-arabinofuranose. It is postulated that the phosphatidyl-inositol acts as a linker to other structural elements of the cell envelope.

Although the basic salient features of the cell wall are well understood it is only in recent years that an understanding of its biosynthesis has emerged, Figure 1.8. There is considerable current research interest in this area, as it is anticipated that the cell wall will contain several chemically unique features offering excellent targets for novel antibiotics.

Figure 1.8: Proposed pathway for the assembly of arabinogalactan, its attachment to peptidoglycan and subsequent mycolyation.



1.5 Biosynthesis of Arabinan

Arabinan is central to the integrity of the mycobacterial cell wall having two important roles to play. In the chemical setting of the mAGP complex, arabinan is an essential structural component of the cell wall framework. Whereas in the context of LAM, arabinan is implicated as being a key surface molecule involved in the pathogenesis of mycobacterial disease. However, it is only in recent years that an understanding of the biosynthesis of arabinan has been obtained.

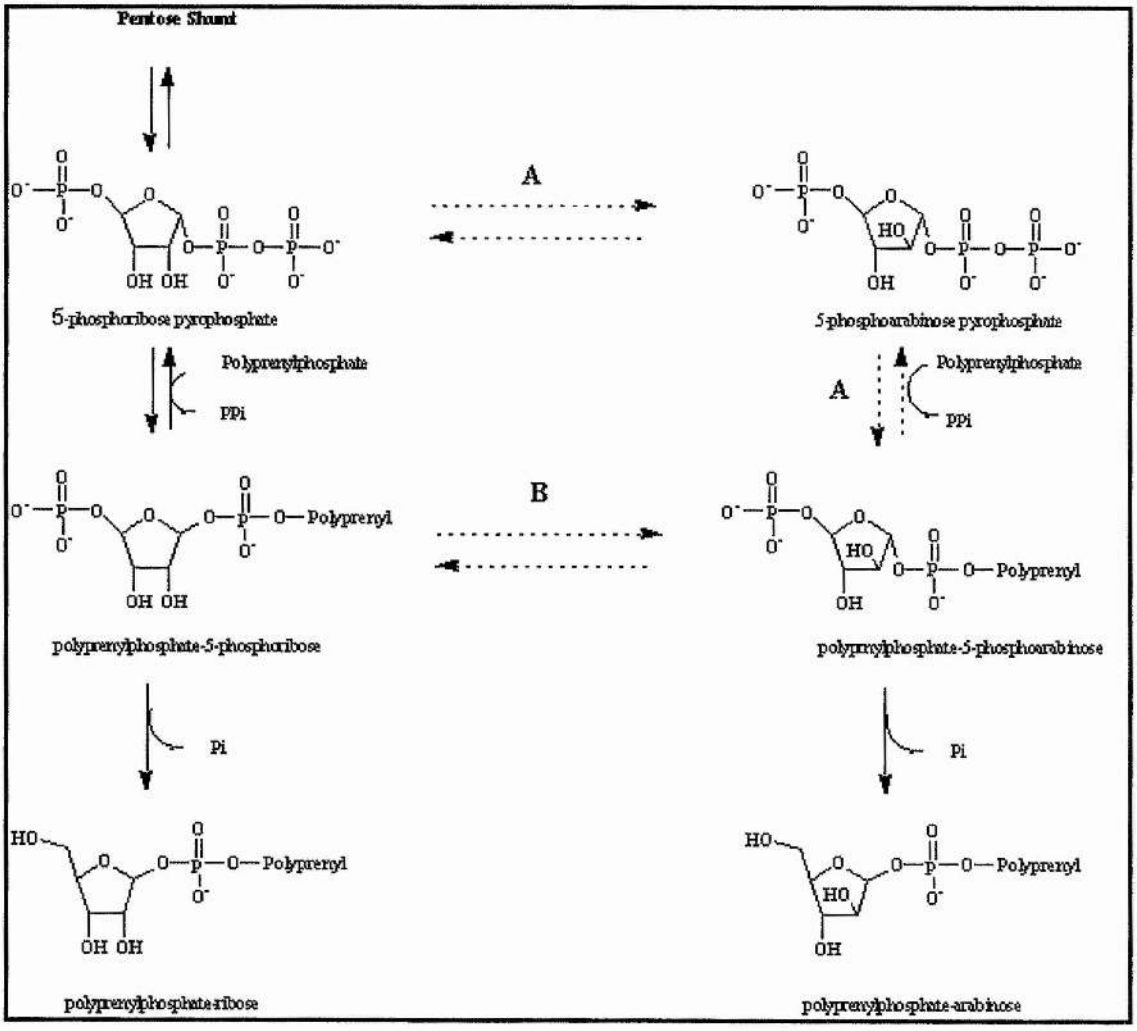
It is postulated that arabinan biosynthesis in mAGP and LAM is via a similar set of arabinosyltransferases, as each structural motif (mAGP and LAM) terminates with the same pentaarabinosyl structure. It follows that as each motif is unique then overall arabinan biosynthesis must either be enzyme regulated, compartmentalised or involve enzymes sufficiently distinct that will allow divergent processing.

Characterisation of lipids extracted from the membrane of *Mycobacterium smegmatis* (*M. smegmatis*) by Woluka and co-workers (1994) suggested that β-D-arabinofuranosyl-1-monophosphodecaprenol (DPA), to be the arabinofuranose precursor in arabinan biosynthesis (Woluka *et al*, 1994). Further work with synthetically derived DPA and a arabinosyl transfer assay demonstrated that DPA was indeed being incorporated into cell wall material (Woluka *et al*, 1994; Lee *et al*, 1995). In addition, if isotopically labelled polyprenylphosphate [^{14}C] arabinofuranose is incubated with particulate enzymes from *M. smegmatis* it is converted into ^{14}C arabinan, and further characterisation reveals ^{14}C hexa-arabinose and ^{14}C di-arabinose oligomers to be present (Schermann *et al*, 1996). Namely the two structural motifs found in mAGP and LAM. As ^{14}C arabinan is detected in both internal and external arrangements of arabinan suggests the donor (DPA) is unique in mycobacteria.

It is currently unclear how DPA is derived biosynthetically. Initially the biosynthesis of DPA was anticipated to involve sugar nucleotides as polyprenylphosphate sugars are generally synthesised via sugar nucleotides (Shibaev *et al*, 1986). However, attempts to identify such analogues failed (Takayama & Kilburn, 1989; Singh & Hogan, 1994). As a result alternative routes were investigated as viable pathways to DPA biosynthesis.

Currently, it is thought that DPA is derived from 5'-phosphoribosyl-1-pyrophosphate (pRpp) via 5'-phosphoarabinosyl-1-pyrophosphate (pApp). Schermann and co-workers (1996) have shown pRpp to be a biosynthetic precursor of polyprenylphosphate-arabinose and that polyprenylphosphate-arabinose is formed via polyprenyl phosphate-5-phosphoarabinose, Figure 1.9. Inhibition of the enzyme forming polyprenylphosphate should inhibit the formation of pApp and subsequently arabinose, thus preventing cell wall synthesis. As D-arabinofurosyl is absent in humans, these enzymes become appealing therapeutic targets, as inhibitors would most likely be non-toxic. If this hypothesis is correct, then the crucial and as yet unresolved step, is the epimerisation from D-ribose to D-arabinose in the formation of pApp. The enzyme involved here would be an inviting subject for structure determination with a view to rational drug design. Alternatively, one might consider inhibiting the biosynthesis of pRpp as a means of ultimately inhibiting arabinan synthesis, but as pRpp is generated by way of the pentose phosphate pathway, inhibition at this level would ultimately inhibit DNA synthesis.

Figure 1.9: The biosynthetic route to polyprenylphosphate-ribose and the proposed route to polyprenylphosphate-arabinose. Established pathways are in unbroken lines and proposed pathways in broken lines (Schermann et al, 1996).



A recent investigation by Draper and co-workers (1997) identified galactosamine (GalN) bound at position 2 of arabinose in the arabinogalactan motifs. As it is present in its free un-N-acetylated state it is postulated that it may help stabilise mycobacterial cell walls through binding of anionic substances. Candidates include the phosphatidylinositol-based LAM or the phosphate linker itself. Similarly as glutamic acid polymers are known to be located within the cell walls of mycobacteria then GalN may bind free carboxyl groups thus

attaching polymers to the cell wall. What is particularly significant about the presence of GalN is that it has only been found in the cell walls of slow growing mycobacteria; it is conspicuously absent in all rapid growing species. It may point to a difference in the cell wall structures of slow and rapid growing mycobacteria and give an insight into why only slow growing species are pathogenic.

The anti-mycobacterial compound ethambutol has been used in the treatment of TB since 1961 (Thomas *et al*, 1961). As well as being a front-line drug recommended for TB treatment it is also used to treat opportunistic infections of AIDS patients caused by the atypical mycobacteria, *Mycobacterium avium.* Several lines of evidence exist to support the theory that the critical target for ethambutol lies in the biosynthesis of arabinogalactan. Mikusova *et al* (1995) demonstrated that addition of ethambutol to cultures of *Mycobacterium smegmatis* resulted in appreciable amounts of DPA accumulation, the proposed arabinan donor in both arabinogalactan and LAM. Secondly, the observations reported by Deng *et al* (1995) and Mikusova *et al* (1995) illustrate that although the *de novo* synthesis of arabinan may be via a similar set of arabinosyltransferases it appears its subsequent polymerisation is through a different route in each structural motif.

The genes encoding ethambutol resistance (*emb A, B* and *R*) have been isolated from *M. avium* and transfected into *M. smegmatis.* The otherwise ethambutol susceptible *M. smegmatis* was rendered resistant to ethambutol after this treatment. Characterisation of the genes revealed that only *embAB* were necessary for an ethambutol resistant phenotype. *EmbR* is expendable though an intact promoter region between *embR* and *embAB* is required. An ethambutol sensitive cell free assay for arabinan biosynthesis showed that overexpression

of the *embAB* is associated with high level ethambutol resistant arabinosyltransferase activity and that *embR* appears to modulate the *in vitro* level of this activity (Belanger *et al*, 1996).

1.6 Biosynthesis of the Linker Region and Galactan

In contrast to the biosynthesis of arabinan, the biosynthesis of the linker region is relatively well understood (Mikusova *et al*, 1996). Pioneering work on the biogenesis of the cell wall of gram positive bacteria (Daffe *et al*, 1990; M^{c}Neil *et al*, 1991) allowed a framework to be established for the biosynthesis of this region of the mycobacterial cell wall. Experiments by Mikusova and co-workers have provided firm evidence that this framework is correct (Mikusova *et al*, 1996).

Radiolabelling of UDP-GlcNAc and subsequent characterisation, including assaying tunicamycin activity, of the two glycolipids (GL_1 & GL_2) formed after incubation with membranes from *M. smegmatis* showed the glycolipids to be polyprenyl based. Similarly, assaying incorporation of radiolabelled TDP-rhamnose into GL_1 and GL_2 showed TDP-rhamnose was only taken up in GL2, suggesting GL2 is in fact polyprenyl-P-P-GlcNAc-rhamnose.

Synthesis of the galactan polymer (alternating 5, 6 linked galactofuranose) has also been confirmed by radiolabelling experiments. Using either UDP [^{14}C]-GlcNAc or UDP-[^{14}C]-Gal_f in incubation with mycobacterial membranes revealing 5-linked Gal_f, 6-linked Gal_f and (5, 6)-linked Gal_f all to be present (Nassau *et al*, 1995). These findings have established that both the linker region and galactofuranose are synthesised on a polyprenyl-phosphate lipid carrier.

If one looks even closer at this biogenesis, and considers the biosynthesis of dTDP-rhamnose and UDP-Gal_f then a plethora of potential targets become available for novel antimycobacterial drug development. The enzymes involved in the biosynthesis of rhamnose and galactan have been identified and are discussed in chapters 6 and 2 respectively. Similarly, as in the biogenesis of arabinan, neither rhamnose nor galactan is utilised in humans, so new drugs should be highly specific and non-toxic. As the mycobacterial linkage unit is the fulcrum of cell wall integrity, it potentially provides a singular site for target-directed chemotherapy, which makes inhibition of this pathway additionally appealing.

1.7 Biosynthesis of Lipoarabinomannan

LAM is not essential for mycobacterial growth in culture. Yet, as it is implicated in the pathogenesis of mycobacterial disease, an understanding of its biosynthesis is essential. It is still unclear how LAM is chemically associated with the cell walls. Extraction methods (boiling crude mycobacterial cell walls in 50% ethanol or Percoll gradient purification) suggest LAM to be firmly, but not covalently, attached to cell walls (Chatterjee, 1997). Furthermore, as whole mycobacterial cells are recognised by monoclonal antibodies to LAM, it implies a portion of the molecule is accessible to the environment (Chatterjee, 1997).

Early work by Brennan and Ballou (1967), established GDP-mannose to be the mannose donor in the initiation of the conversion of phosphatidylinositol (PI) to PI-$mannoside_x$ where X represents the number of added mannoses.

Radiolabelling experiments showed the synthesis of a novel α (1,6)-linked linear form of lipomannan, and the concomitant decrease in C_{35}/C_{50}-P-mannoses. As the synthesis is susceptible to amphomycin, an inhibitor of polyprenyl-P requiring translocases, this suggests

the C_{35}/C_{50}-P-mannoses are the mannosyl donors rather than UDP-mannose, as synthesis of the simpler PIM_2 was not inhibited by amphomycin. The source of the mannose residues is now established as being from UDP-mannose in the biosynthesis of the simpler PIM's and from the polyprenyl-mannoses in the elongation of the mannose backbone. The origin of the α(1, 2) mannose branches is not currently known but is thought to be from GDP-mannose (Besra *et al*, 1997).

1.8 Biosynthesis of the Mycolic Acids

Mycolic acids, being unique to mycobacteria, are the major determinants of the impregnability of the mycobacterial cell wall. Structurally, the long chain α-alky-β-hydroxyfatty acids are complimented by various glycolipids including α, α'-trehalose dymycolate (TDM), and the analogous monomycolate (TMM). It appears mycolate acids enable the cell to survive the hostile environment of the macrophage by forming an impermeable lipid bilayer (Liu *et al*, 1995). The biosynthetic pathway of the mycolic acids is understood to be the position of attack for several anti-TB drugs including the front-line drug Isoniazid (Banerjee *et al*, 1994).

The biosynthesis of mycolic acid is proposed to be via a modification of normal fatty acid metabolism where short chain fatty acids are extended and altered to form lipids of the required length (George *et al*, 1995).

Research to identify the enzymes and genes responsible for mycolic acid deposition is ongoing in many independent laboratories, and to date has resulted in the identification of several targets on which to base the rational design of new anti-mycobacterial drugs.

Belisle and co-workers (1997) have identified a mycolyl transferase that catalyses the transesterification of non-radioactive mycolic acids to [^{14}C] α, α'-trehalose, yielding both [^{14}C] TMM and [^{14}C] TDM. The enzyme responsible for this exchange has been purified and analysed by 2D-PAGE to reveal two proteins of identical pI and similar molecular weights, 31 and 34kDa, furthermore N-terminal sequencing of each protein yields an identical sequence (RPGLPVEY). A comparison of this sequence with that of the antigen 85 (Ag85) complex (Ag85 A, B and C) from *M. tuberculosis* reveals a striking similarity (FSRPGLPVAY). Ag85 is a triad of gene products implicated in mycobacterial pathogenesis through their fibronecting binding capabilities (Peake *et al*, 1993). Investigations with the individual components of the Ag85 complex confirmed mycolytransferase activity. Belisle and co-workers (1997) went on to demonstrate each component of the Ag85 complex to show carboxylesterase activity. Interestingly, transesterification of mycolic acids dictates carboxylesterase activity. An analysis of the primary sequences of Ag85 A, B and C disclosed the consensus Gly-x-Ser-x-Gly that is conserved among other fatty acyl transferases. X-ray crystallography has shown serine to be the active residue of the catalytic triad serine, aspartic acid/glutamic acid and histidine (Kazlauskas, 1994). Site-directed mutagenesis (serine to alanine) destroyed the carboxylesterase activity to confirm the role of serine in the Ag85 component.

Given that the mycolyl transferase identified to catalyse transesterification of mycolic acids is reactive to monoclonal antibodies specific for the Ag85 complex further suggests the members of Ag85 complex have a role to play in mycolic acid metabolism (Belisle *et al*, 1997).

A screen of putative competitive inhibitors of this metabolism identified 6-azide-6-deoxy-α,α'-trehalose (ADT) as being able to suppress growth of *Mycobacterium aurum* (an established surrogate of *M. tuberculosis*) but not of *Escherichia coli*. Along with this observation an additional unknown compound was accumulated on inhibition, presumably an intermediate in the mechanistics of the reaction. That ADT inhibits synthesis of TMM and TDM and subsequently cell wall synthesis, demonstrates that mycolate transfer is essential in cell wall integrity, identifying these enzymes (Ag85 A, B and C) as viable targets.

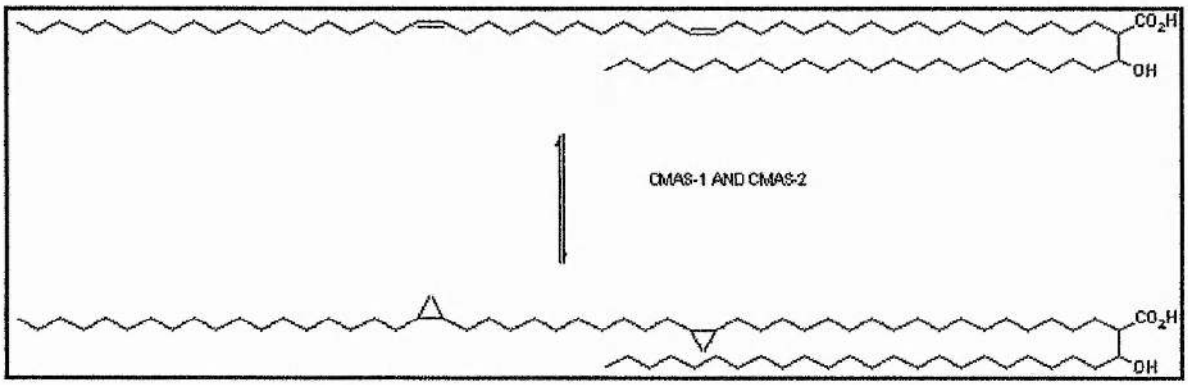
The major mycolic acids produced by *M. tuberculosis* are defined as the α-mycolic acids which can be further divided into two subclasses α_1 and α_2. Each class averages some 78-80 carbons in length (Danielson and Gray, 1982). α_1, the most abundant of the class, contains two *cis* unsaturated bonds (proximal and distal) which are open to cycloprotonation Figure 1.10. Cycloprotonation is significantly more common in slow growing mycobacteria than the more rapid growing saprophytic species (Minniken, 1982). In addition, modification at the distal position has been shown to confer on the slow growing bacteria, a greater resistance to H_2O_2. Furthermore, differential scanning microcalorimetry has shown cycloprotonation at the proximal position to raise the observed transition temperature by 3°C (George *et al*, 1995). Together these results imply that cycloprotonation of mycolic acids may be an important adaption of slow growing pathogenic mycobacteria.

Cycloprotonation at each position is catalysed by two independent enzymes, CMAS-1 and CMAS-2, and the genes responsible for expression of these enzymes have been identified (George *et al*, 1995). It is postulated that *S*-adenosyl-L-methionine is an enzyme cofactor, as analysis of the primary sequences identifies a consensus motif (V/L) LDXGXGXG which

has been previously been proposed to play a role in binding of this co-factor (Ingrosso *et al*, 1989).

The two enzymes help make up a family of four proteins that catalyse the transfer of a methylene group from *S*-adenosyl-L-methionine to an unsaturated lipid. The others are from the other mycobacterial pathogen *M. leprae* and *E. coli.* As no three-dimensional structures of these enzymes are available no firm molecular mechanistic details exist as to how these enzymes function. Mammals do not cycloprotonate unsaturated lipids, making these enzymes appealing therapeutic targets for novel drug intervention.

Figure 1.10. : Cycloprotonation of the major mycolic acids in M. tuberculosis.



Isoniazid, one of the front-line anti-tuberculosis drugs, has been used to treat tuberculosis since 1952 (Middlebrook, 1952). It is only in the last decade that an understanding of both its mechanism of action and the development of resistance to it has become apparent. Isoniazid is a pro-drug activated by catalase peroxidase upon administration to its active form. Interestingly, the gene encoding catalase peroxidase *(katG)* is found to have deletions or missense mutations in around 20-30% of isoniazid resistant cases of *M. tuberculosis* (Zhang *et al*, 1992).

The target for activated isoniazid has been shown to be mycolic acid biosynthesis (Quemard, 1991). Two open reading frames (*orf's*) from an isoniazid resistant mutant from *M. smegmatis* have been identified. Subcloning studies showed the second of these open reading frames to be sufficient in conferring the isoniazid resistance phenotype and was duly labelled *inhA*. This differed from the wild-type (*wt*) gene (non-isoniazid resistant) by a single base pair mutation, T to G, resulting in substitution of alanine 44 for serine 44. Allele exchange experiments confirmed drug sensitivity was directly related to the single mutation (Banerjee *et al*, 1994). Further characterisation revealed the *wt* protein to bind NADP and catalyse the reversible reaction of 2 – *trans* long chain enoyl (C_{16}-C_{20})-acyl carrier proteins (ACP'S). More specifically, by following the oxidation of NADH spectrophotometrically, Dessen and co-workers (1995) showed the gene catalysed the reduction of 2-*trans*-octenoyl-ACP, confirming that *inhA* is involved in mycolic acid biosynthesis. Kinetic studies by the same group suggest the role NADH plays in the mechanism to be crucial. K_m and V_{max} of the mutant protein were found to be similar to the *wt* for the reduction. However, the Michaelis constant for NADH was five times higher in the mutant, Table 1.2, implying the interactions between the enzyme and co-factor may be related to the mechanism of drug resistance (Dessen *et al*, 1995).

Table 1.2: Kinetic evaluation of the wt and S94A mutant inhA catalysed reduction of 2-trans-octenyl-ACP by NADH (Dessen et al , 1995).

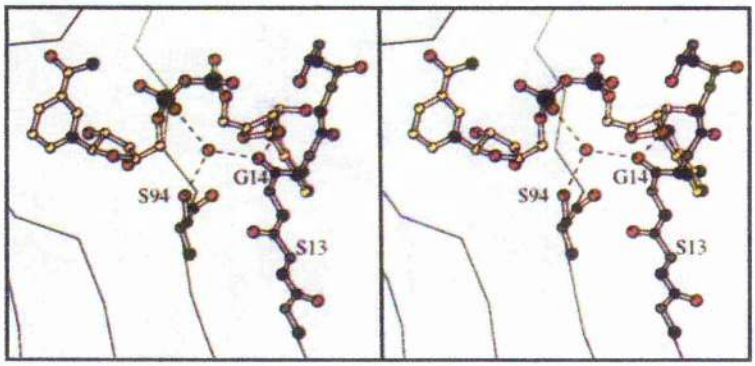
Enzyme	K_m (μM)		V_{max}
	2-*trans*-octenoyl	NADH	(μmol min^{-1} mg^{-1})
Wild-type	2 ± 1	8 ± 0	2.2 ± 0.4
S94A	3 ± 1	38 ± 2	3.1 ± 0.6

To investigate these interactions further, Dessen and co-workers (1995) determined the three-dimensional structures of both the wild type and the mutant proteins. The structures are described as having an overall chairlike appearance, composed of seven β strands and eight α-helices, Figure 1.11. The 'legs' and 'seat' of the protein are in the orientation typically known as the Rossman fold (Rossman, 1975), and it is here the cofactor binds to the protein.

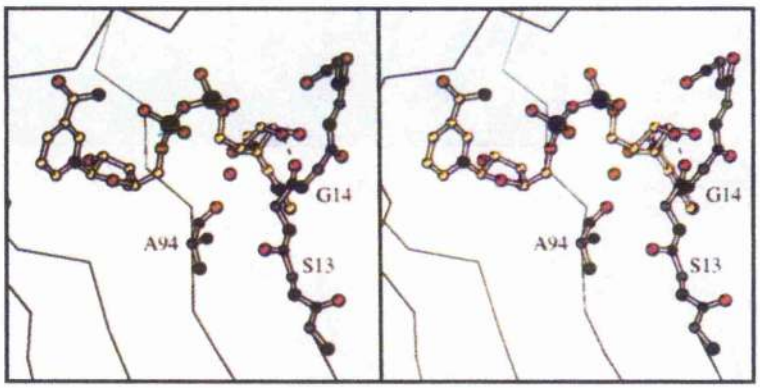
Figure 1.11: Schematic ribbon representation of InhA complexed with one NADH molecule (blue) (Dessen et al, 1995).

It is in the binding of NADH that the structures differ significantly, Figure 1.12. NADH is stabilised in the wild type structure by thirteen hydrogen bonds involving a combination of polar amino acids, backbone atoms and six water molecules. One of these hydrogen bonds, involving a bridging water molecule, is disrupted in the S94A structure, Figure 1.12b. In the wild type, a hydrogen bond exists between O of the P_N phosphate of NADH and a well ordered water molecule. This water molecule in turn forms a hydrogen bond with the hydroxyl of serine 94, the main chain O of glycine 14 and two other main chain atoms (alanine 22 and isoleucine 21). In the mutant structure, glycine 14 faces the opposite direction, away from the NADH, curtailing the extent of hydrogen bond between NADH and the protein. Overall the disruption of the hydrogen bonding accounts for around two to three kcal mol^{-1} of binding energy which will explain the mutants five-fold reduced affinity for NADH evident from kinetic studies.

Figure 1.12:a) Stereo view of the atoms involved in the binding of NADH in a) wild type and b) mutant InhA. N = blue, O = red and phosphate = green. Carbons atoms are yellow in NADH and grey in protein (Dessen et al, 1995).



b)

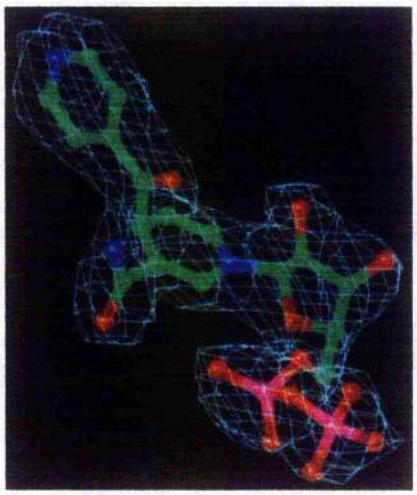


Dessen *et al* (1995) proposed that the disruption of the hydrogen bond network associated with co-factor binding is directly related to the subsequently acquired resistance of the mutant. They suggest protein / isoniazid binding is inhibited either by the reformation of a nucleotide inhibitor complex or upon conformational changes in the enzyme active site when NADH binds, which alters the substrate binding site.

Rozwarski and co-workers (1998) have elucidated the crystal structure of the activated form of isoniazid in complex with NADH and *inh*A.

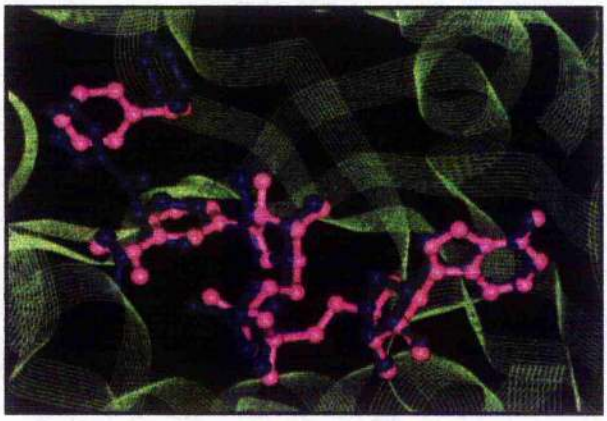
Crystals were grown from an incubation mixture containing *inhA*, NADH, isoniazid and $MnCl_2$ and resulted in the activated form of isoniazid associating with *inhA* bound NADH. The structure revealed an extended NADH molecule. The isonicotinic acyl group is attached through its carbonyl carbon to the carbon at position 4 of the nicotinamide ring of NADH. The carbon at position 4 of the nicotinamide ring of the isonicotinic acyl-NADH complex is tetrahedral, retaining one hydrogen, Figure 1.13.

Figure 1.13: Isonicotinic acyl-NADH superimposed onto the $2F_o$-F_c electron density map contoured at 1σ (C = green, O = red, N = blue and P =magenta) (Rozwarski et al, 1998).



A comparison of the overall protein structures of the *inhA* / NADH and *inhA* / isonicotinic acyl-NADH shows the structures to be almost identical. The only significant difference in the protein is located in the position of the side chain of phenylalanine 149, Figure 1.14. In the binding of isonicotinic acyl-NADH the bulky side chain is rotated ~ 90° to form π-π interactions with the pyridine ring of the isonicotinic acyl group.

Figure 1.14: Superposition of the active sites of inhA with bound NADH (pink) and with bound isonicotinic acyl-NADH (blue) (Rozwarski et al, 1998).



Previous kinetic isotope experiments (Quemard, 1995) with *inhA* have demonstrated a preference for NADH binding prior to the acyl-ACP substrates which would result in the majority of the wild type enzyme being NADH bound, making it susceptible to isoniazid attack. Once isoniazid is bound, the enzyme is effectively inactivated and mycolic acid biosynthesis inhibited. Thus the vulnerable enzyme population resting with NADH bound is substantially reduced and enzyme that is actively turning over is resistant to isoniazid.

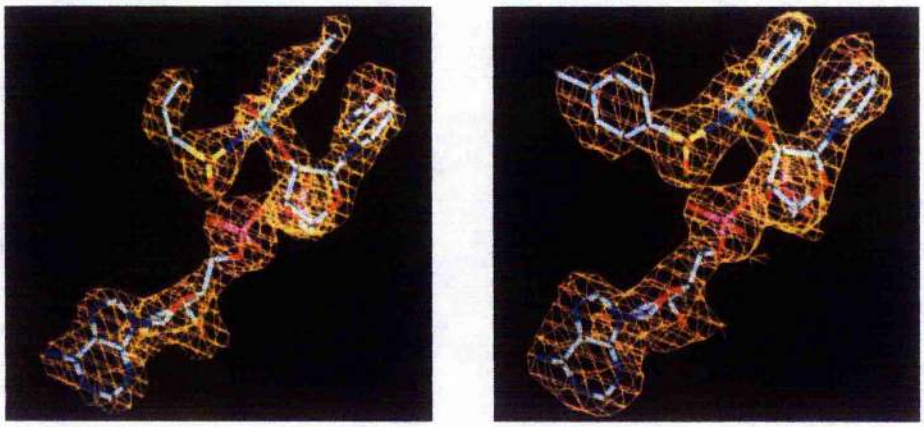
The existence of this complex (isonicotinic acyl NADH bound *inh*A) illustrates the crucial role NADH plays in the mechanism of action of isoniazid. Quite simply its presence is essential for isoniazid activity.

The predicted sequence of *inh*A shows strong sequence similarity with *E. coli* Enoyl Reductase (ENR) and the *Brassica napus* analogue which are known to catalyse the last reductive step of fatty acid biosynthesis (Baldock *et al*, 1996). NAD^+ is an essential co-factor in this reaction and is required for binding a known class of inhibitor of the reaction, the diazoborines (Baldock *et al*, 1998). In ENR the single mutation glycine 93 to serine 93 is responsible for inducing resistance to the diazoborines by conferring a reduced activity for NAD^+, and consequently the inhibitor. Baldock and co-workers (1996) have determined the three-dimensional structure of native ENR and of two ENR / inhibitor complexes. The native ENR structure (2.1Å) was solved by multiple isomorphous replacement and used as a model in molecular replacement in order to solve the two complexed structures (2.2 and 2.5Å). The native structure is tetrameric. Each monomer is composed of a 7 stranded parallel β-sheet flanked by 5 α helices. An additional helix lies along the top of the β-sheet. The overall topology is again reminiscent of the Rossmann fold, as is expected with a

NADH binding enzyme. Each monomer contains a loop of 10 disordered residues opposite the NADH binding site which become ordered on inhibitor binding.

Both diazoborines bind in a closely related manner. The bicyclic rings of the diazoborines lie adjacent to the nicotinamide ring of NAD^+ allowing the formation of extensive π stacking interactions. The major difference is in the positioning of the respective tosyl and propyl moieties, Figure 1.15. The tosyl moiety lies perpendicular to the bicyclic ring system and is stabilised by a combination of interactions between nearby main and side chain atoms. The propyl group however, folds back on the ring system in a manner not unlike a scorpion's tail. It also forms interactions with the main chain and side chains of nearby residues.

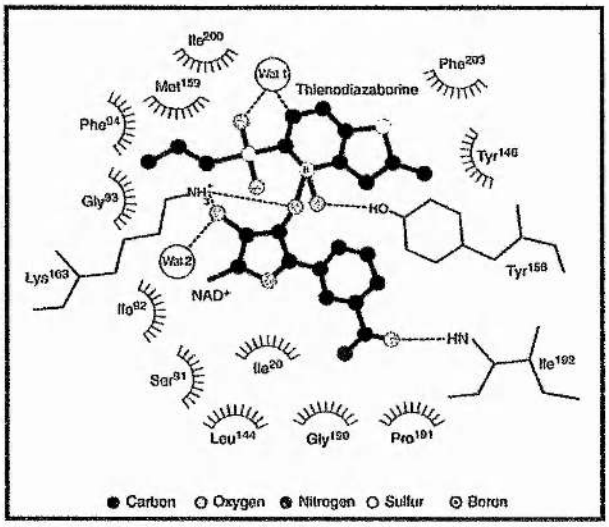
Figure 1.15: Adopted conformation of the diazoboranes in the two complexes in E. coli ENR diazoborane complexes (Baldock et al, 1998).



The most significant interaction comes between the boron atom of diazoborane and the 2' hydroxyl of nicotinamide. At a distance of 1.7Å it is comparable with a boron to oxygen covalent bond length (1.6Å). The boron atom is at the centre of a tetrahedral array, as is required of a boron atom forming four covalent bonds, Figure 1.16. The loss of biological activity upon replacement of the B-N group by C-C in diazoborines (Grassberger *et al*, 1984)

is readily explained by this observation, that is the covalent bonding forming potential is removed. Boron has been observed to play a similar role in the inhibition of serine proteases by boronic acids where it also forms covalently bound tetrahedral adducts, resulting in chemical modification of the active site serine leading to enzyme inactivity (Zhong *et al*, 1991).

Figure 1.6 : Schematic representation of the interactions made by the NAD$^+$ thienodiazoborane with the enzyme and water molecules. Hydrogen bonds are dashed lines and Van der Waals interactions are hydrophobic arcs (Baldock et al, 1998).



The analogies between the Baldock *E. coli* structures and the *M. tuberculosis* structures reported by Dessen and Rozwarski are clear. Both enzymes demonstrate a requirement for a dinucleotide co-factor, bound at a conserved 'dinucleotide binding site', described by Rossmann (1975). It is on the inhibition of these enzymes however that the most significant similarities arise. Each enzyme is inhibited through the binding of their respective inhibitors in a covalently bound tetrahedral adduct via the associated co-factor. That is, for enzyme inhibition prior cofactor binding is essential. Removal of each co-factor induces the enzymes

resistant to the inactivating drugs. A single point mutation is all that is required in each enzyme to bring this about.

Other reactive inhibitors involving NAD(P) have been described (Taillefumier *et al*, 1996) but none act with such potency. The differences in such structures originate from steric interactions preventing the formation of covalent bonds of the nature described by Baldock. Indeed, further studies on the ENR enzyme show that substitution of glycine 93 for amino acids with small side chains for example, serine, alanine, cysteine and valine, hardly affected the catalytic parameters of the enzyme, but still had the effect of inducing resistance to the drug. Substitutions with even larger amino acids for example, arginine, histidine, lysine and glutamine rendered the bacterial host completely resistant to the drug (de Boer *et al*, 1999).

The findings of these studies involving the *M. tuberculosis* analogue enzyme ENR suggest the type of linkage that may be required not only in new anti-mycobacterial drugs but any inhibitor that involves NAD(P) as a co-factor. However, the underlying problem in the formation of these bonds is a steric one. If this can be overcome then a new class of antibacterial agents may be uncovered. In relation to the mycobacterial cell wall, these drugs would be targeted at inhibiting mycolic acid biosynthesis, overcoming the problems of MDR-strains of TB, particularly isoniazid resistant strains. However, given the potential toxicity of boron, such inhibitors will require boron replacement with an alternative atom.

Although the structures described by Baldock and Rozwarski both involve the formation of bisubstrate analogues between a diazoborane and NAD+ or isoniazid and NADH respectively, the mechanism of drug action and resistance inferred are quite different. The isonicotinic acyl NADH / *inhA* structure provides a basis for the design of agents that inhibit *inhA*. Yet what remains unanswered, is if any drug designed to act in a mode similar to

isoniazid will be efficacious against MDR-TB, including isoniazid resistant strains. *InhA* is undoubtedly a target for novel drug intervention, but one feels different modes of inhibition from that displayed by isoniazid will be required if novel anti-mycobacterial drugs are to be developed and to be efficient against MDR-TB strains, particularly isoniazid strains. The diazoboranes inhibit the *E. coli* analogous enzyme ENR, and should be assayed for activity in the *inhA* system. If effective they would provide a basis on which to design new anti-mycobacterial drugs. As mentioned previously, the toxicity of boron is another problem which still requires to be overcome.

1.9 Concluding Remarks

With more people estimated to die from TB in the forthcoming year than any other in living history, the urgency with which new anti-mycobacterial agents are required has never been greater. However, increased awareness of the TB problem, as a result of WHO intervention, has resulted in several co-ordinated research efforts world-wide aimed at tackling the problem. As a direct consequence of this a greater understanding of the mycobacterial cell wall has emerged. The basic and salient structural features are fully established and current research is now directed at dissecting the biosynthetic pathways involved in the assembly of these molecules. The cell wall structure of mycobacterium is unique and its biosynthesis reveals suitable targets for intervention, many of which are discussed in this and forthcoming chapters. One particularly appealing enzyme is UDP-galactopyranose mutase, which catalyses the formation of galactofuranose in the mAGP complex. This target, along with the others discussed may lead to new classes of drugs for the treatment of re-emerging mycobacterial disease.

Chapter 2

Initiating a crystallographic study of UDP-galactopyranose mutase from *Escherichia coli.*

2.1 Summary

Galactofuranose, the thermodynamically disfavoured form of galactose, is a key component of mycobacterial cell walls where the interconversion of UDP-galactopyranose and UDP-galactofuranose is catalysed by UDP-galactopyranose mutase. UDP-galactopyranose mutase, from *Escherichia coli*, has been crystallised in a form suitable for X-ray diffraction studies. Two crystal forms have been obtained. A single crystal of dimensions 0.4 x 0.2 x 0.05mm was grown over one week from a solution of 10mg ml^{-1} protein. The yellow crystal diffracted to 2.7Å when fresh but decayed due to X-ray damage in less than 10 hours. The data were indexed in an orthorhombic space group (P222, $P2_12_12_1$, $P2_12_12$ or $P222_1$) with cell dimensions a = 56 Å, b = 94 Å, c = 134 Å, $\alpha = \beta = \gamma = 90°$. This crystal form has not been reproduced. However, crystals of both native and selenomethionine protein, grown over 5 days from a solution of 8mg ml^{-1} protein in hanging drops have been consistently reproducible. Crystals grow as thin plates of dimensions 0.04 x 0.02 x ~0.005mm and multiple rounds of seeding are required to obtain crystals (0.4 x 0.2 x ~0.05mm) large enough for X-ray diffraction studies. Once at maximum size the crystals diffract to beyond 3.0Å on BM 14 at the European Synchrotron Radiation Facility (E.S.R.F.), France. Crystals are monoclinic, space group $P2_1$ and unit cell dimensions a = 71.12 Å b = 58.42 Å c = 96.38 Å, $\beta = 96.38°$. A 92% complete native data set, redundancy 2.5, R_{merge} 5.0% has been collected at the E.S.R.F. to 2.9Å. The Matthew's co-efficient is $2.35Å^{-3}$, equivalent to 47% solvent content for a dimer in the asymmetric unit. Data sets on potential derivatives and the selenomethionine variant protein have all been collected at the same source.

Crystals of UDP-galactopyranose mutase are high in mosaic spread, the value of which varied from crystal to crystal but was always above 1°. Even the largest crystals obtained by

seeding diffracted weakly and the resolution observed varied considerably from crystal to crystal. In-house data collection, particularly with the selenomethionine crystals, has been limited by the weak diffraction and low resolution limit of the data. This prevented characterisation of the mosaic spread in-house. Only with the intense radiation available on BM14 has it been possible to fully characterise the mosaicity problem.

Crystals of UDP-galactopyranose mutase have also been obtained from a different organism, *Klebsiella pneumonia.* The crystals are currently too small for X-ray diffraction, however crystallisation conditions have yet to be fully optimised.

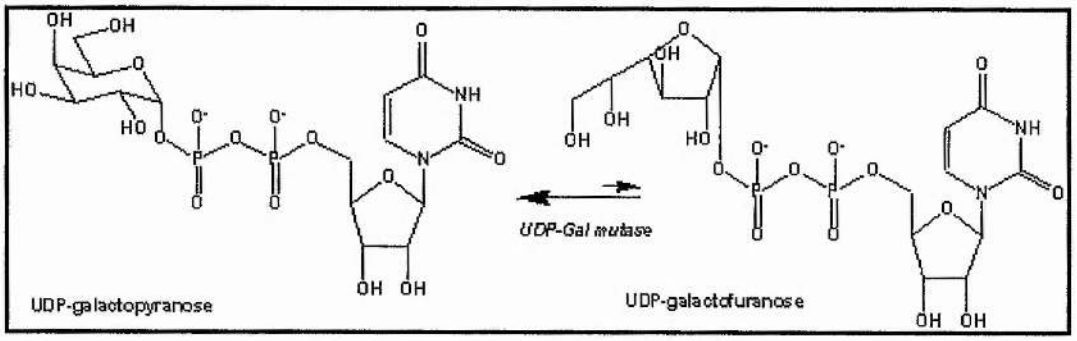
2.2 Introduction

Mycobacterial Cell Wall

Galactofuranose, the thermodynamically unfavourable form of galactose is not commonly found in nature. It is however, found in various macromolecules including bacterial O antigens, fungal exopolysaccharides, glycoproteins and the cell wall of mycobacteria (Stevenson *et al*, 1994; Latge *et al,* 1994; Nassau *et al*, 1996 and Moraes *et al*, 1987).

The recent structure determination of the O antigen of *Escherichia coli* K-12 together with the elucidation of the sequence of its *rfb* gene cluster (Stevenson *et al*, 1994) has allowed identification of the gene labelled *orf6* that encodes for the enzyme UDP-galactopyranose mutase (mutase) (Nassau *et al*, 1996). The enzyme catalyses the ring interconversion of UDP-galactopyranose (UDP-Gal_p) and UDP-galactofuranose (UDP-Gal_f), Figure 2.1.

Figure 2.1: Interconversion of UDP-galactopyranose and UDP-galactofuranose



Gal_f is not found in humans and is structurally unique. It is therefore an appealing therapeutic target as inhibition of its biosynthesis would assist the development of new anti-mycobacterial drugs that are likely to be highly effective and non-toxic.

UDP-galactopyranose Mutase from E.coli.

E. coli mutase was the first enzyme from any organism shown to catalyze the interconversion of UDP-Gal$_p$ and UDP-Gal$_f$, though the enzyme has also been isolated from *Klebsiella pneumoniae* (Koplin *et al, 1997)* and *M. tuberculosis* (Weston *et al,* 1998). Mutase from each source is known to incorporate flavin adenine dinucleotide (FAD). That the enzyme incorporates FAD and possibly NAD(P) is not surprising as analysis of the primary sequence identifies a consensus β-α-β (Rossmann fold) motif consistent with nucleotide binding enzymes (Bellamacina, 1996). The Rossmann Fold, so called after it was first identified by Michael Rossmann in 1975, is a conserved region among dinucleotide binding proteins. In mutase the sequence is located close to the N terminal incorporating amino acids 3 to 39. This sequence from *E.coli* mutase was used as the subject of a Basic Logical Alignment Search Tool (BLAST) search at the Structural Classification of Proteins (SCOP) server, (http://scop.mrc-lmb.cam.ac.uk/scop/) where its compared to the primary sequences of proteins of known tertiary structure. The search identified several positive matches where the sequence was located in other proteins, where either FAD or NAD(P) was a bound co-factor. An example is shown in Figure 2.2.

Figure 2.2: Output from a BLAST search performed with SCOP. The sequences display 51% identity and 61% homology. The conserved amino acids are crucial in the binding of the associated dinucleotide.

YDYIIVGSGLFGAVCANELKKLNKKVLVIEK	Mutase
YDYII+GSG A E LN KV +IE+	
YDYII I GSGGAAFSSAI EAVALNAKVAMIER	M.I.R.

M.I.R. is mercuric ion reductase from *Bacillus sp.* strain RC607, (Schiering, 1991) and is known to contain a FAD / NAD(P) binding domain.

Subjecting the complete mutase sequence to a BLAST search reveals no other known protein with a greater sequence identity than 30%, other than alternative versions of mutase (date of search 14/07/99). Unsurprisingly, the region (80 amino acids) of glutathione reductase (EC 1.6.4.2) from *Haemophilus influenzae* where this identity is found contains a nucleotide binding motif (Fleischmann, 1995).

UDP-galactopyranose Mutase from Klebsiella pneumoniae

Mutase is also encoded for by one of six genes recently isolated that are involved in the synthesis of D-galactan. D-galactan, $[\rightarrow 3)$-β-D-Gal_f-$(1\rightarrow 3)$-α-D-Gal_p-$(1\rightarrow]$, forms the backbone of the O - side chain polysaccharide in the lipopolysaccharide of *K. pneumoniae*. The enzyme shows approximately 40% sequence identity to the protein from *E.coli*. Analysis of the primary sequence identifies a βαβ ADP binding domain at the N-terminus consistent with what is found in the *E.coli* enzyme (Koplin *et al*, 1997). As the protein also displays a distinct yellow colour it is anticipated this region would be involved in FAD binding. As in the case of the *E.coli* enzyme, enhanced activities are also observed when the protein from *K. pneumonia* is incubated with NADH or NADPH together with FAD (Koplin *et al*, 1997). As the reaction involved does not involve the net transfer of electrons, the mechanistic role of these cofactors is unknown. Our work with *E.coli* mutase demonstrates FAD incorporation and suggests the enzyme also requires NAD(P).

2.3 Experimental

Overexpression of E. coli BL21 (DE3) (pORF6).

All materials used were obtained from Sigma (Poole, UK) unless otherwise stated. Prior to all cultivation procedures, all containers, pipette tips and media were sterilised by autoclaving at 120°C for 30 minutes.

In a typical experiment five millilitres of Miller's Luria Broth Base (LB) medium containing ampicillin (200μg ml^{-1}) was inoculated with a single colony of BL21 (DE3)(pORF6) and incubated overnight at 37°C. This culture was used to inoculate 500ml of similarly treated LB medium and the resultant solution shaken at 200rpm, 37°C. The optical density was monitored until the A_{600} reached 0.6 whereupon IPTG was added to a final concentration of 1mM. The incubation was continued for a further 3 hours under the same conditions. The cells were then harvested by centrifugation (10000g, 4°C, 30 minutes) and frozen at -78°C until required.

Preparation of Cell Extracts.

Frozen cell paste was thawed in ice-cold lysis buffer (50mM TRIS pH 7.6, 1mM DTT, 1mM PMSF) at a concentration of 0.5g wet cell paste for each mL buffer. Lysozyme (200μg ml^{-1}), DNase (20μg ml^{-1}) and EDTA (2mM) were added and the solution left at room temperature for 30minutes before the cell contents were released by sonication (140kW, 20kHz) at 4°C for 6 x 30 seconds, resting on ice in between. Cell debris was pelleted by centrifugation (20000g, 4°C, 30 minutes). The cell extract was transferred to a clean tube and the contents assayed by the Bradford assay (Bradford, 1976) and SDS-PAGE.

Overexpression of mutase in LB by this protocol typically yielded 1mg soluble protein for every litre of media. The majority of the protein was insoluble and seen in the lysis precipitate when viewed on denaturing gels. Alteration of the expression temperature and the IPTG concentration to try and overcome this problem had little effect, Table 2.1.

Table 2.1: Different expression conditions for overexpression of UDP-galactopyranose mutase in LB and yield of protein obtained.

Media	Temperature (°C)	IPTG (mM)	Yield (mg L^{-1})
LB	37	1	1.0
LB	37	0.4	0.9
LB	30	1	1.2
LB	30	0.4	1.3
LB	25	1	1.1
LB	25	0.4	1.1

As no assay was readily available to detect mutase's activity purification from inclusion bodies was not considered as a means of purification. The harsh agents required to release the protein from inclusion bodies, that is guanidine HCl or urea, would denature the protein and on removal of these agents correct refolding of the protein could not be assessed. With this in mind alternative expression media were sought to improve the level of soluble protein expressed. A total of 5 different media were tried at both 30°C and 37°C and at 0.4 and 1mM IPTG concentrations, Table 2.2. The media that resulted in the optimum amount of soluble protein was terrific broth (Maniatis, 1982). The yield of soluble protein increased to

some 8.1mgL^{-1} by growing the cells to an optimum cell density at 37° and then on addition of 1mM IPTG lowering the temperature to 30°C and leaving the cells for a further 3 hours.

Table 2.2: Different media tried and yields of soluble protein obtained after purification.

Media	IPTG (mM)		IPTG (mM)	
	0.4	1	0.4	1
Post Induction Temperature	30°C		37°C	
LB[a] + Sorbitol	1.1	1.3	1.2	1.3
LB + Betaine	1.0	1.2	1.1	1.1
LB + Sorbitol/Betaine	1.1	1.2	0.9	1.0
M9[b]	1.1	1.3	1.1	1.1
Terrific Broth[c]	7.7	8.1	2.2	2.4

[a] To 950 ml of deionized water add 10g bacto-tryptone, 5g bacto-yeast extract and 10g Na Cl. Adjust pH to 7.0

[b] To 750ml of deionized water add 200ml 5 x M9 salts (see below), 2ml 1M $MgSO_4$, 2ml glycerol, 0.1ml $CaCl_2$.
5 x M9Salts: To 150ml of deionized water add 12.8g $Na_2HPO_4.7H_2O$, 3g KH_2PO_4, 0.5g NaCl and 1g NH_4Cl. Make to 200ml after dissolution of solutes and sterilise separately.

[c] To 900ml of deionized water add 12g bacto-tryptone, 24g bacto-yeast extract and 4ml glycerol. Sterilised separately, to 90ml of deionized water add 2.31g KH_2PO_4, 12.54g K_2HPO_4. Adjust the volume to 100ml after the salts have dissolved and add to the sterile media.

Preparation of Selenomethionine Variant UDP-galactopyranose Mutase.

As increasing numbers of protein structures are being solved by Multi-wavelength Anomalous Dispersion (MAD) then different methods for the overexpression of seleno-methionine substituted protein are becoming apparent. Three different routes are outlined in this thesis and each was evaluated in terms of the yield of soluble protein and convenience.

One method of overexpressing selenomethionine protein involves using a cell line auxotrophic for methionine. The advantage of this being that maximum selenium incorporation in the protein is more likely than the alternative (using a non-auxotrophic

strain). However, unless the gene of interest is already in an auxotrophic cell line, some DNA manipulation will be required to make the cells producing the protein auxotrophic for methionine.

DNA Extraction and Transformation into the Methionine Auxotroph B834

Single colonies of *E. coli* BL21 (DE3) (pORF6) were selected from agar/ampicillin plates and used to inoculate 5ml LB/ampicillin (200μg ml^{-1}) broth and incubated overnight at 37°C with shaking. 1.5ml of this culture was transferred to a sterile 1.5ml Eppendorf tube and microcentrifuged for 2 minutes at room temperature. Supernatant was discarded and the pellet resuspended in 100μl 50mM glucose, 10mM EDTA, 4mg ml^{-1} lysozyme, 100μg ml^{-1} Rnase A, 25mM TRIS pH 8 in sterile H_2O. This solution was left at room temperature for 10 minutes whereupon 200μl of fresh sterile SDS was added and the tube shaken thoroughly before being incubated on ice for a further 5 minutes. 150μl of ice cold 3M NaOAc pH 4.8 was added and the solution mixed well and incubated at 4°C for a further 45 minutes. The samples were then centrifuged for 10 minutes before carefully transferring 400μl of the supernatant into a clean sterile Eppendorf tube, taking care not to transfer any of the precipitate. 400μl of a 1:1 mixture of phenol and chloroform was added and the solution vortexed and centrifuged for 2 minutes before transferring the top aqueous phase to a clean tube. Ice cold ethanol was added with thorough mixing and the solution centrifuged again for 10 minutes. The supernatant was carefully drained by pipette and tube inversion. The DNA pellet was redissolved in 100μl of 0.1M NaOAc pH 6 and incubated at room temperature for 15 minutes. 200μl absolute ethanol was added and mixed well. This solution was centrifuged for 10 minutes. The supernatant was drained as before and the pellet air-dried for 45 minutes before re-suspending in 50μl sterile H_2O. The DNA was

stored at 20°C until required. DNA content (when stained with ethidium bromide, which intercalates DNA) was quantified by running on an agarose gel against markers of known composition and comparing band intensity under UV light. Generally for a high copy number plasmid (Marie-France Giraud, personal communication) as used here, pET-11a, 0.5 - 2μg DNA is sufficient for successful transformation with efficient electrocompetent cells.

Aliquots (20μl) of B834 electrocompetent cells were thawed on ice before addition of 1μg and 10μg of mutase DNA, along with positive and negative controls. These suspensions were transferred to electroporation cuvettes and electroporated for 5 $x10^{-3}$ seconds at 2500 volts. After electroporation 1ml of sterile LB medium was added and the cells left to recover at 37°C for 1 hour. The cells were then microcentrifuged for 2 minutes and 800μl of the supernatant removed. The cells were resuspended in the remaining media and streaked on LB agar/ampicillin plates and incubated overnight at 37°C.

Approximately twice as many colonies grew with the 1μg transformation compared to the 10μg transformation. The positive control yielded thousands of colonies indicating the cells were indeed highly competent and the negative control no colonies, confirming no contamination took place in the course of the experiment.

Overexpression and Purification of B834 (pORF6) Mutase in LeMaster Medium.

Single colonies of B834 (pORF6) were selected from an agar/ampicillin plate and used to inoculate 2 x 5ml of LeMaster medium, Table 2.3, supplemented with

1mg L^{-1} thiamine. This starter culture was incubated at 37°C overnight with agitation and used to inoculate 2 x 500ml of LeMaster medium. Incubation continued at 37°C, 200rpm before addition of 1mM IPTG at an OD_{600} of 0.6. The temperature was lowered to 30°C and

the cells left shaking for a further 16 hours before being harvested by centrifugation (10000*g*, 4°C, 15 minutes).

Purification of seleno-methionine variant protein was performed in the same manner as the native with the only difference being an increased presence of DTT. 15mM DTT as opposed to 1mM was used throughout purification to ensure there was no oxidation of selenomethionine. In all cases buffers were thoroughly degassed before use. Protein was eluted at the same position on the NaCl gradient (anion exchange) but the yield was 10 fold down on the native and the distinguishing yellow colour associated with FAD was absent. The increased concentration of DTT appeared to be stripping FAD from the protein. Subsequently the DTT concentration was dropped to 2mM and purification duly proceeded as the native. Protein quantity was assayed by Bradford and protein homogeneity by SDS-PAGE. Typical yield of protein was only an eighth of that obtained after expression in terrific broth at $1mgL^{-1}$.

Overexpression of B834 (pORF6) Mutase in Defined Medium.

A single colony of B834 (pORF6) was selected from an agar/ampicillin plate and used to inoculate 5ml of LB medium. This culture was incubated at 37°C overnight with agitation. The cells were harvested the following morning by centrifugation and the supernatant discarded. Cell pellets were resuspended and washed in 10ml of the defined medium, Table 2.3, and again harvested by centrifugation. This process was repeated three times to minimise the carry over of methionine to the larger cultures. Cells washed in defined medium were used to inoculate 1000ml of similar media; incubation was continued at 37°C, 200rpm. The OD_{600} was monitored until it reached 0.6 AU when the temperature was

lowered to 30°C and 1mM IPTG added. Incubation was continued for a further 16 hours before the cells were harvested by centrifugation (10000*g*, 4°C, 15 minutes).

Purification of the modified protein was as outlined previously for Le Masters medium. Protein quantity was assayed by Bradford and protein homogeneity by SDS-PAGE. 1.2mg of selenomethionine variant protein was obtained from 1 litre of media.

Overexpression of BL21 (DE3) (pORF6) Mutase by Methionine Pathway Inhibition.

As an alternative to the methionine auxotroph one can produce selenomethionine protein by inhibiting methionine biosynthesis throughout overexpression. Some amino acids (isoleucine, lysine and threonine) are known to block methionine biosynthesis in *E. coli* when added to minimal media at a high enough concentration (Doublie, 1997). Under these conditions, selenomethionine is freely incorporated by the protein during overexpression.

A single colony of BL21 (DE3) (pORF6) was selected from an agar/ampicillin plate and used to inoculate 5ml of LB medium. This culture was incubated at 37°C overnight with agitation. The cells were harvested the following morning by centrifugation and the supernatant discarded. Cell pellets were resuspended in M9 minimal media, Table 2.4, before being added to 1000ml of fresh M9 media. Cells were grown to an OD of 0.6 at A_{600}. At this point the amino acids lysine, phenylalanine and threonine were added to a final concentration of 100mg L^{-1} and isoleucine, leucine and valine to 50mg L^{-1}. Selenomethionine, 60mg L^{-1}, was added at this stage. The media was left to shake at 37°C for 15 minutes before 1mM IPTG was added and induction continued for 16 hours at 30°C. Purification was as outlined for Le Masters medium. Only 0.9mg of homogenous selenomethionine protein was obtained after purification.

As the yield of protein obtained was essentially the same from each media, the methionine inhibition pathway was initially used to overexpress selenomethionine protein as it was the most convenient. However, characterisation of this protein by mass spectrometry suggested selenium incorporation was not complete and for this reason a switch was made to the methionine auxotroph route for expression of the selenomethionine protein. As the yield of soluble protein obtained from each media was essentially the same, the defined medium was chosen as it was the most convenient.

Table 2.3: Composition of Le Masters and defined media.

Le Masters Media	Substance[c]	Defined Media
0.5[a]	Alanine	0.045
0.58[a]	Arginine	0.045
0.4[a]	Aspartic Acid	0.045
0.03[a]	Cystine	0.045
0.67[a]	Glutamic Acid	0.045
0.33[a]	Glutamine	0.045
0.54[a]	Glycine	0.045
0.06[a]	Histidine	0.045
0.23[a]	Valine	0.045
0.23[a]	Isoleucine	0.045
0.23[a]	Leucine	0.045
0.23[a]	Threonine	0.045
0.42[a]	Lysine	0.045
0.13[a]	Phenylalanine	0.045
0.1[a]	Proline	0.045
2.08[a]	Serine	0.045
0.17[a]	Tyrosine	0.045
0.05[e]	**Selenomethionine**	**0.1[d]**
0.5[a]	Adenine	0.5
0.67[a]	Guanosine	0.5
0.17[a]	Thymine	0.5

Table 2.3 (cont): Composition of Le Masters media and defined media.

0.5[a]	Uracil	0.5
1.5[a]	Succinic Acid	-----
1.5[a]	Sodium Acetate	-----
0.75[a]	Ammonium Chloride	1
2.16[a]	Sodium Hydroxide	-----
10.5[a]	Di-Potassium Phospate	-----
	Sodium Chloride	0.5
-----	Potassium di-Hydrogen Phosphate	3
-----	Thiamine	0.002
10[b]	Glucose	5
0.25[b]	Magnesium Hepta Sulphate	0.025
0.0042[b]	Iron Hepta Sulphate	0.004
8.4µl[b]	Conc. Sulphuric Acid	-----

[a] Made up in 900ml of dH_2O and autoclaved.
[b] Made up in 100ml of dH_2O and 0.22µm filtered. Added to cool autoclaved portion of media ([a]).
[c] All amino acids are the L-enantiomer.
[d] Added after the media has been 0.22µm filtered. [e] Added at time of induction.
[f] All compounds mixed overnight at room temperature in 1000ml dH_2O. 0.22µm filter sterilised.

Table 2.4: Composition of M9 Minimal media.

To 750ml sterile dH_2O add:	mL
5 x M9 Salts[a]	200
1M Magnesium Sulphate[b]	2
20 %Glucose[c]	20
1M Calcium Chloride[b]	0.1

[a] Dissolve 64g $Na_2HPO_4.7H_2O$, 15g KH_2PO_4, 2.5g NaCl and 5g NH_4Cl in dH_2O and autoclave.
[b] Prepare $MgSO_4$ and $CaCl_2$ separately, autoclave and add after diluting 5 x M9salts to 1L.
[c] Sterilised by 0.22µm filtration

Purification of Recombinant UDP-galactopyranose mutase

In planning the purification protocol, the objective is to obtain as maximum a level of purity as is efficiently possible. This is achieved by exploiting the different properties of the protein such that the number of steps required for sufficient purification is minimised. Techniques are evaluated for capacity, protein recovery, resolving power and cost effectiveness when deciding what methods to utilise and in what order they should be performed.

A concern in the early stages of purification is to reduce the working volume of sample; hence a high capacity technique is sought where sample concentration will be achieved. Typically, precipitation is the first step in a purification strategy as it is high capacity and almost all the protein is recovered; in addition the working volume of sample is greatly reduced. Although precipitation has these advantages the purification achieved can be low, hence subsequent steps place more emphasis on the resolving power of the technique. Separation based on charge, hydrophobicity, pI and affinity are more likely candidates at this stage as each have high resolving powers. A combination of two or more of these methods are often required together, exploiting different properties of the protein to achieve the desired state of homogeneity.

In designing the purification protocol for mutase all of the outlined factors were taken into consideration. The first step was ammonium sulphate fractionation. This removes fouling lipids and hydrophilic proteins from the sample.

The theoretically determined pI of mutase is 6.93. This is the point of no net charge of the protein. At pH values above this point the protein will become more positively charged and

hence be suited to anion exchange chromatography. To determine the optimum pH for separation a series of separations were performed across the pH range 6 – 9. At pH levels below 7 the eluted protein was stripped of FAD. Above pH 7 the best resolution was obtained at pH 7.6. To compliment this separation a similar experiment was performed with hydrophobic interaction chromatography; again optimum separation was found to be at pH7.6. The purification protocol described here is equally applicable to protein expressed in either LB or terrific broth.

Ammonium Sulphate Fractionation and Anion Exchange Chromatography.

Cell extract was prepared as previously outlined and brought to 10% $(NH_4)_2SO_4$ saturation by the slow addition of 5.6g $(NH_4)_2SO_4$ per 100ml cell extract with gentle stirring at 4°C. This solution was centrifuged to remove precipitated contaminants and the supernatant transferred to a clean tube. Supernatant was dialysed 3 times against a 1000ml solution of 50mM TRIS pH7.6, 1mM DTT to remove ammonium sulphate before application to an anion exchange chromatography column. Dialysed protein was 0.02μm filtered (Acrodisk) and applied to a 7.84ml anion exchange column (Perseptive Biosystems, HQ) equilibrated in buffer A (50mM TRIS pH 7.6, 1mM DTT) in 100mg aliquots at 5ml min^{-1}. Protein was eluted was over 15 column volumes (CV's) by application of a 0-500mM NaCl gradient, Figure 2.3. Protein elution was monitored on-line at 280 and 450nm. Mutase was easily identified by its distinct yellow colour, a function of the FAD co-factor. Appropriate A_{450} peaks from subsequent runs were pooled and concentrated under nitrogen pressure in an Amicon ultrafiltration cell.

Hydrophobic Interaction Chromatography.

Pooled fractions from the anion exchange column were brought to 40% NH_4 $(SO_4)_2$ saturation by the slow addition of 2.4g $(NH_4)_2SO_4$ per ml at 4°C with gentle stirring. Protein was again 0.02μm filtered before being applied at 5ml min^{-1} to a 7.84ml hydrophobic interaction chromatography column (Perseptive Biosystems, HP) equilibrated in buffer B (50mM TRIS, 40% $(NH_4)_2$ SO_4, 1mM DTT). Protein was eluted in a decreasing $(NH_4)_2SO_4$ gradient (100-0%) over 15 CV's, Figure 2.4. Appropriate A_{450} fractions were again combined and concentrated as before. Protein concentration was determined by the Bradford method (Bradford, 1976). Purity was assessed by SDS-PAGE, Figure 2.5, and IEF PAGE.

Figure 2.3: Anion exchange chromatography trace. The A_{450} trace shadows the A_{280} helping to identify mutase.

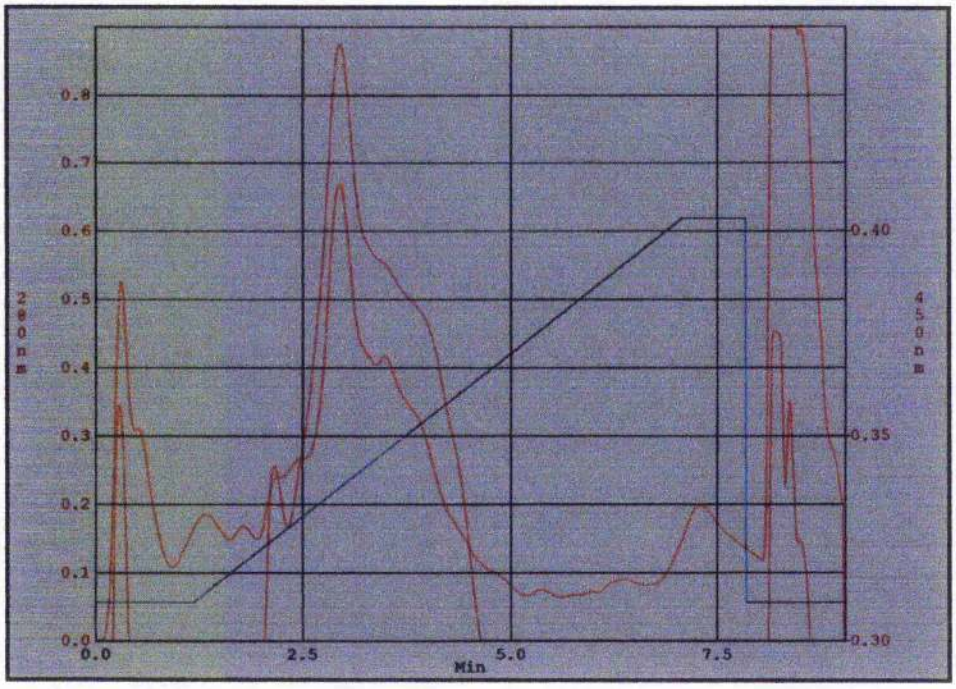


Figure 2.4: Hydrophobic interaction chromatography trace. Again the A_{450} trace shadows the A_{280} identifying mutase.

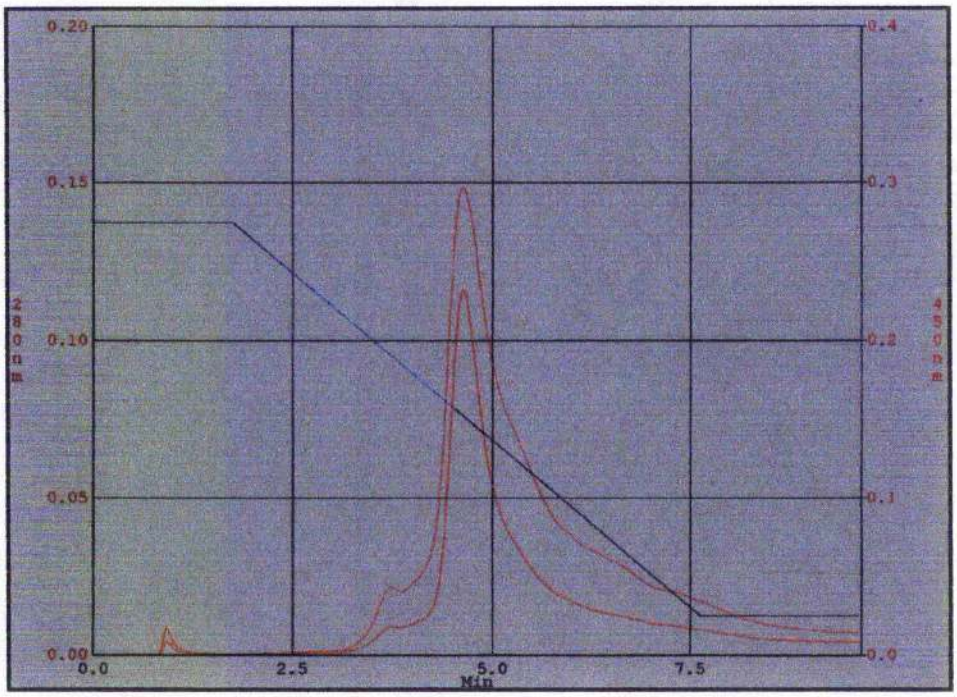
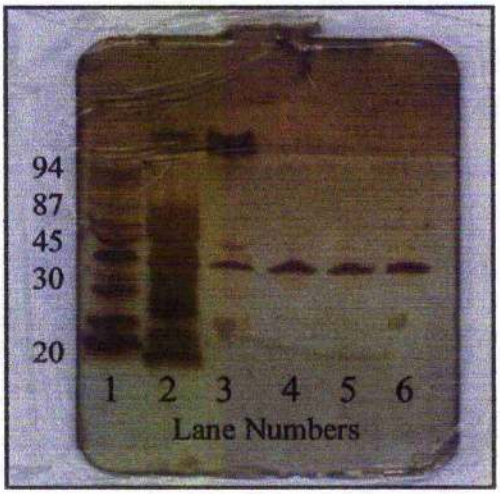


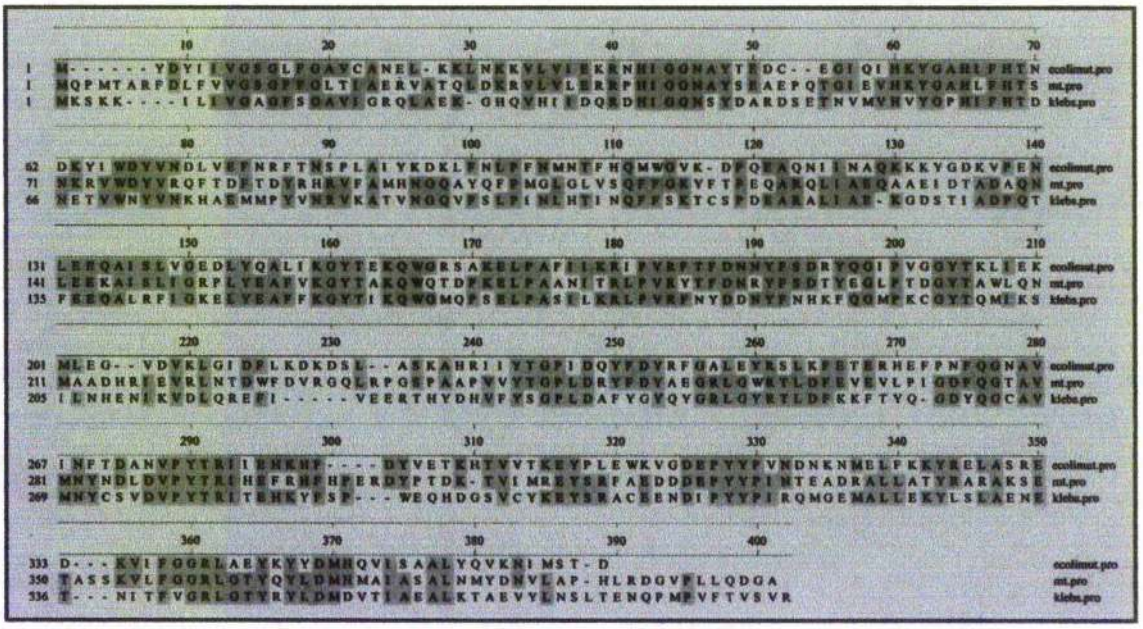
Figure 2.5: Silver stained SDS-PAGE displaying protein purity after subsequent purification steps.
Lane 1: Molecular markers (kDa)
Lane 2: Crude UDP-galactopyranose mutase.
Lane 3: Post HQ UDP-galactopyranose mutase.
Lanes 4, 5 & 6: Post HP UDP-galactopyranose mutase.



N-Terminal Sequencing of Native Protein.

100µg of purified protein was dialysed in H_2O in preparation for N-terminal sequencing by Edman degradation (Stryer, 1988). 5 cycles were performed and the resulting sequence was unambiguous. The first 6 residues were found to be MYDYII, which exactly match the first 6 residues found in the predicted sequence for *E. coli* mutase, Figure 2.6. The N-terminal sequencing revealed no heterogeneity.

Figure 2.6: Alignment displaying the sequence identity between mutase from E. coli, M. tuberculosis and K. pneumonia. The shaded area represents the amino acids common to each enzyme. Sequence identity is 40%.

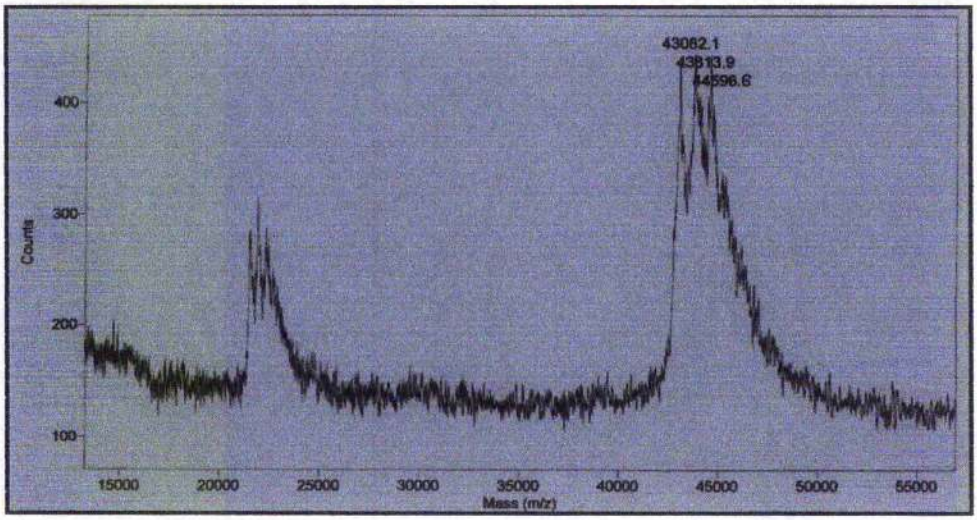


Analysis of Native Protein by Mass Spectrometry.

As isolated, three very closely related peaks are evident by mass spectrometry, Figure 2.7. The reasons for this are not exactly clear and several hypotheses may be made. Three peaks are visible in the molecular weight region predicted for mutase at 43082.1, 43813.9 and

44596.6 Da. The theoretical molecular weight predicted for mutase is 42960 ± 8 Da (Nassau *et al*, 1996).

Figure 2.7: Analysis of native UDP-galactopyranose mutase by mass spectrometry.



A gas phase reaction may occur after laser vaporisation in which FAD binds to the protein. This would create a molecule of molecular weight 43789.5 Da, only 24.4 Da different from the observed middle peak from mass spectrometry. Similarly if NADP is also incorporated the theoretical peak would be at 44554.9, which again is only 41.1 Da away from the heaviest peak predicted from mass spectrometry. The peaks either side of the middle peak are 783 Da higher and 731 Da lower in predicted weight. As the molecular weight of NADP is 765Da, a scenario where NADP is and is not incorporated may be apparent, Table 2.5. It should be remembered that the enzyme from *K. pneumonia* has also been shown to incorporate NADP (Koplin *et al*, 1997), hence its presence here should not be surprising.

Table 2.5 : A comparison of molecular weights as predicted theoretically and by mass spectrometry.

Theoretical Molecular Weight	*Native* *42960*	*+ FAD* *43789*	*+ FAD + NADP* *44554*
Molecular Weight by Mass Spec.	*43082*	*43813*	*44596*
Molecular Weight by Mass Spec *(Assuming FAD* Incorporated, i.e. *43813)*	*- NADP* *43048*		*+ NADP* *44578*

To try and clarify the uncertainty arising from this analysis the experiments will be repeated with different matrices.

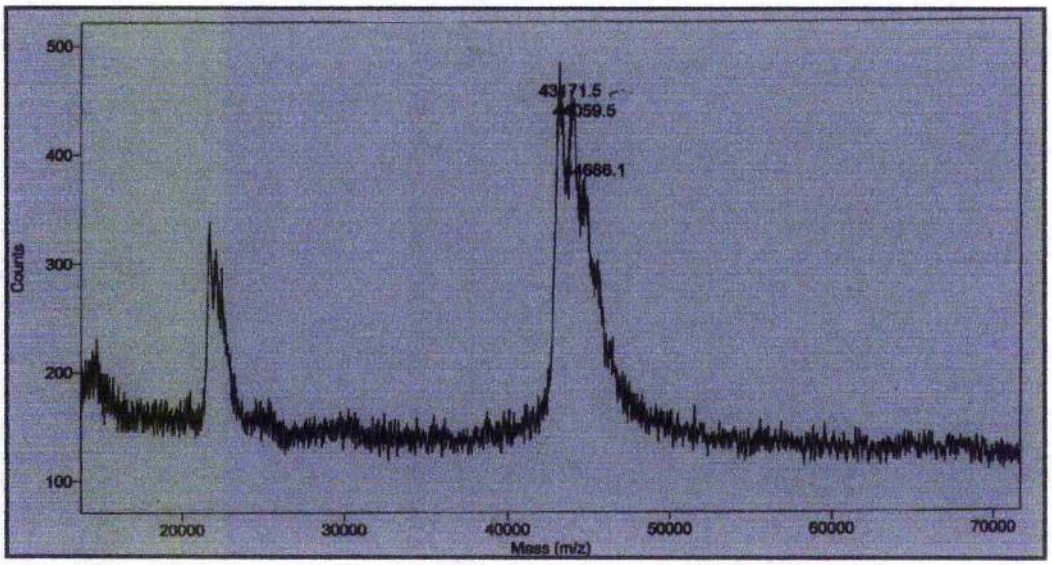
Analysis of Selenomethionine Variant Protein by Mass Spectrometry.

Analysis of selenomethionine substituted protein by mass spectrometry revealed the same breakdown pattern as the native protein. Three peaks were again evident in the region expected for mutase with no other peaks being observed, Figure 2.8. Although selenium incorporation is not complete at an average of 70-80%, it was thought to be of a high enough level to proceed with crystallisation. The shortfall in selenium incorporation is not surprising given that this sample originated from a 'methionine inhibition' overexpression experiment. A higher level of incorporation was anticipated from the protein produced using the auxotrophic expression system. Table 2.6 shows the predicted molecular weights from the native mass spectrometric analysis assuming 5, 6, and 7 methionines have been substituted in comparison to the experimentally found values.

Table 2.6: A comparison of molecular weights calculated theoretically and found experimentally by mass spectrometry for selenomethionine protein..

	Molecular Weight (Da) as Predicted by Native Mass Spec. +			
	5	***6***	***7***	***methionine subs***
	44038	*44083*	*44128*	
Molecular Weight Predictions from Mass Spectrometry.	*43171*	*44059*	*44686*	

Figure 2.8: Mass spectrometric analysis of seleno-methionine variant mutase



Dynamic Light Scattering With Protein Expressed in LB.

For the successful crystallisation of a macromolecule it is generally accepted that a monodisperse solution of protein is preferable (Ferre - D'Amare and Burley, 1996). Whilst common protein characterisation techniques, electrophoresis and chromatography, may detect contaminated protein solutions they do not reveal any aggregation that may be present within (http://www.protein-solutions.com). There have been several extensive studies to assess the use of dynamic light scattering in determining the crystallisability of protein solutions, (Mikol *et al*, 1990; Zulauf and D'Arcy, 1992; and D'Arcy, 1994). The general

consensus arising is one of agreement in that monodisperse macromolecular solutions will crystallise, whereas aggregated solutions or polydisperse systems rarely yield crystals. Dynamic light scattering (DLS) can be used quickly to analyse a protein solution for its monodispersity and subsequently its suitability for crystallisation. A simplified outline of the theory involved is presented below.

The translational diffusion coefficient of a macromolecule undergoing Brownian motion can be determined by exposing the solution to a wavelength of 780nm and collecting the light scattered from these molecules. The light is collected by a lens and transmitted through fiber optics to a photodetector where the photons are converted to electrical impulses. This data allows calculation of the hydrodynamic radius. As the technique is sensitive to aggregation an estimate of the sample polydispersity can calculated, as the intensity of light scattered is proportional to the square of the mass of the protein. If the polydispersity expressed as a percentage of the hydrodynamic radii is on average less than 30%, then the protein will likely crystallise (http://www.protein-solutions.com).

Purified mutase from LB was concentrated to 10mg ml^{-1} under nitrogen pressure (Amicon) and dialysed 3 times against 20mM TRIS pH 7.6, 2mM DTT in preparation for analysis by dynamic light scattering. The instrumentation used for light scattering was the Dynapro 801 (Protein Solutions). The optical cell was equilibrated by injecting a 0.02μm filtered solution of 20mM TRIS pH 7.6 until the photon count rate dropped below 10. On attempting to inject the protein through a 0.02μm filter the resistance felt was too great and the experiment aborted. That such resistance was encountered suggests the protein solution is highly aggregated and unsuited to crystallisation.

Crystallisation of UDP-galactopyranose Mutase From LB Broth.

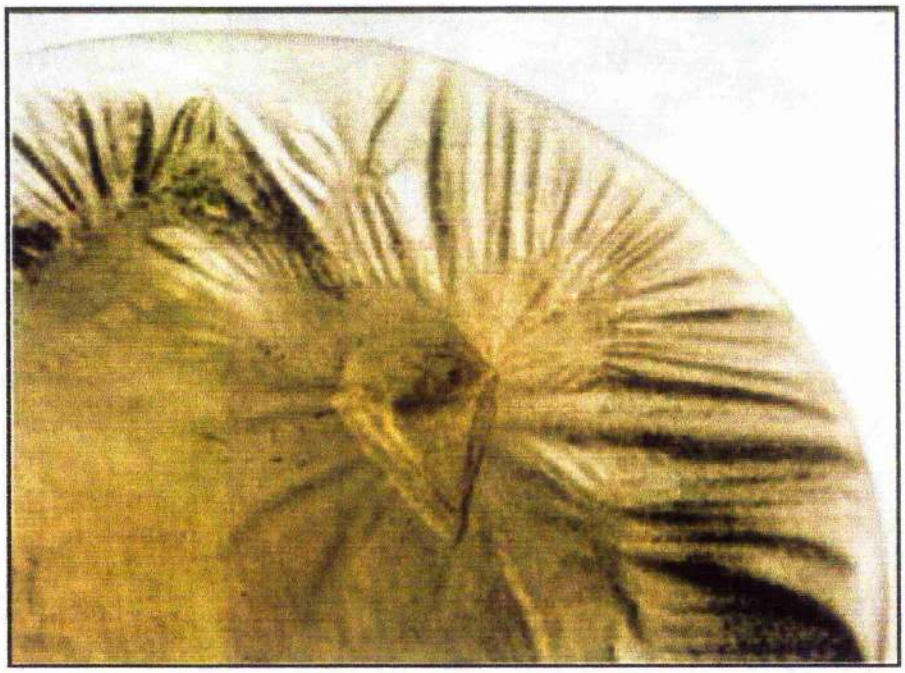
Mutase from LB(and thereafter terrific broth) concentrated to 10mg ml^{-1} was used for initial crystallisation condotions. PEG 6K, PEG 6K/LiCl, $(NH_4)_2SO_4$, MPD and NaCl grid screens where precipitant concentrations are varied against pH were used to search for initial crystallisation conditions. The hanging drop method was employed throughout all crystallisation trials (Ducruix and Giege, 1992) where 2µl of protein is mixed with an equal volume of precipitant on a siliconized glass cover slip. The cover slip was inverted and the drop left to equilibrate against a known precipitant in a sealed environment. Each grid screen was set-up at room temperature and left to equilibrate at a constant temperature of 20.5°C. After 4 days one yellow crystal of dimensions 0.4 x 0.2 x 0.1mm was grown from a solution of 10mg ml^{-1} protein, 10% PEG 6K (w/v), HEPES pH 7.0, Figure 2.9. Table 2.7 charts the different crystallisation parameters tried to reproduce the crystallisation.

Table 2.7: Grid screens of each listed precipitant were performed at room temperature in a search for initial crystallisation conditions. 10% PEG 6K (w/v), 0.1M HEPES pH 7.0 was the only condition that resulted in a crystal.

Precipitant / Concentration (w/v) / (v/v)	**Buffer / pH 0.1M**	**Additive**	**Protein Concentration mg ml^{-1}**	**Mixing[a]**	**Crystals**
Peg 6K		---			Yes
Peg 6K / LiCl	4, 5, 6.	---			No
$(NH_4)_2SO_4$	7, 8 & 9.	---	6 & 10	Yes	No
MPD		---			No
NaCl		---			No

[a] On addition of an equal volume of mother liquor to the protein drop this solution can be mixed by drawing back on the pipette and redispensing the mixed solution.

Figure 2.9: Yellow crystal of UDP-galactopyranose mutase from E.coli.



Unfortunately this crystallisation could not be reproduced. Extensive efforts, utilising exactly the same stock solution as precipitant and a multitude of similar conditions, Table 2.8, proved fruitless. Precisely why this crystallisation could not be reproduced is unclear. One reason would be the multitude of isoforms that exist when the protein is expressed in LB. Separating the protein by charge as opposed to mass disclosed several very closely related species not detectable by SDS-PAGE. Crystallisation of such species may prove problematic and difficult to reproduce, as indeed is discovered here. To assess if the problem of reproducibility lay in the method of overexpression protein was 're-expressed' in the original media, LB. Unfortunately this did not resolve the problem. Batch variation is another possibility, however no subsequent batch characterised by conventional techniques (SDS-PAGE, IEF and DLS) disclosed any difference from the original sample with the

exception that subsequent batches were more homogenous. The failure to reproduce this crystallisation resulted in having to re-screen for 'initial' crystallisation conditions.

Table 2.8: Different crystallisation parameters tried in attempts to reproduce the initial crystallisation of UDP-galactopyranose mutase.

Precipitant **Concentration (w/v) / (v/v)**	**Buffer / pH 0.1M**	**Additive**	**Protein** **Concentration mg ml^{-1}**	**Mixing**	**Crystals**
PEG 6K / 5 – 30%	Citric Acid 5.0-5.9 MES 6.0 – 6.9 HEPES 7.0-8.0	---	3, 6, 8,10 & 12	Yes & No	No
PEG 6K / LiCl 5-30%	HEPES 7.0-8.0		6, 10 &18	Yes & No	No
PEG 200 PEG 600 PEG 1000 PEG 3000 PEG 4000 PEG 8000 PEG 20 000	MES 6.0 – 6.9 HEPES 7.0-8.0 TRIS 8.1 – 9.0	No	6 & 12	Yes & No	No
PEG MME 2K 15 – 25% PEG MME 5K 15 – 25%	HEPES 7.0-8.0	No	4, 8, & 12	Yes	No
PEG 6K	HEPES 7.0	Hampton[a*] Add I & Add II.	10	Yes and No	No

[a] Hampton Research Additive Kits I & II for Crystallisation.
* Full Compilation listed in Appendix II

Dynamic Light Scattering and Crystallisation of Mutase From Terrific Broth.

The protein obtained after expression in terrific broth was more abundant and homogenous (a single IEF species) than that obtained from LB. With these factors in mind, together with the inability to reproduce the initial crystallisation results from LB, a screen for crystallisation conditions from protein expressed in terrific broth was performed.

A protein sample and the Dynapro 801 were prepared as before for a dynamic light scattering experiment. On injection of the 0.02μm filtered protein the photon count rate was monitored until it reached a stable reading. Measurement then commenced and the results are displayed in Table 2.9.

Table 2.9: Dynamic light scattering results for UDP-galactopyranose mutase. The average polydispersity value as a function of the hydrodynamic radius is 15.6%.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDa	Temp. C	Count Rate	Base Line	Poly D%
1	0.871	530	3.7	---	73	17.9	217	1.000	---
2	0.860	544	3.6	---	67	17.4	213	0.999	---
3	0.873	534	3.7	---	69	17.6	211	0.999	---
4	0.884	546	3.6	---	66	17.4	206	1.000	---
5	0.874	538	3.6	0.601	68	17.7	204	0.999	16.6
6	0.876	543	3.6	---	68	17.6	200	1.000	---
7	0.873	547	3.6	---	68	17.1	198	1.000	---
8	0.873	534	3.6	---	66	17.1	207	1.000	---
9	0.868	526	3.7	---	68	17.5	205	0.999	---
10	0.867	538	3.7	0.544	70	17.4	209	1.000	14.7

The expression and purification had been refined such that the protein appeared considerably more suitable for crystallisation. The results from light scattering indicate the protein solution to be monodisperse and it was anticipated that any success in crystallisation now would be more reproducible. However, despite the improved route to crystallisation, the original crystallisation was still not reproducible. As a result commercially available screening kits from Hampton Research and an additional screen containing a further 119 unique conditions[a] (John MacLean, personal communication) were screened for initial crystallisation conditions using the hanging drop method as previously outlined. The results are illustrated in Table 2.10.

Table 2.10: Screening for crystallisation conditions of UDP-galactopyranose mutase with commercially available crystallisation kits.

Precipitant **Concentration (w/v) / (v/v)**	**Buffer / pH** **0.1M**	**Additive**	**Protein** **Concentration** **mg ml^{-1}**	**Mixing**	**Crystals**
Hampton CS I[a]			6, 12 & 18	Yes	Hampton I # 41
Hampton CS II[a]		No			No
J Mac[a]					No
PEG 4K 20% Isopropanol 12%	**HEPES 7.5**		**8**	**Yes**	**Yes**
				No	**No**

[a] Full Compilation listed in Appendix I.

The crystallisation of the enzyme was found to be readily reproducible using the conditions quoted in Table 2.10, although not in a form suited to analysis by x-ray diffraction. However, this is often the case with the first crystals obtained and optimisation of these parameters may yield crystals of higher quality. Increasing the PEG concentrations in 1% increments between 10% and 30%, (all other parameters as before), and similarly varying the isopropanol concentration between 5 and 20%, found the optimal conditions for crystal growth over a range of protein concentrations to be 20% PEG 4K and 12% isopropanol. The optimum protein concentration for crystal growth is 8mg ml^{-1}. Table 2.11 lists these experiments.

Table 2.11: Initial optimisation of primary hit with Hampton Screen I, solution number 41.

Precipitant **Concentration** **(w/v) / (v/v)**	**Buffer / pH 0.1M**	**Additive**	**Protein** **Concentration** **mg ml^{-1}**	**Mixing**	**Best Crystals**
PEG 4K 10 – 30% Isopropanol 10%	HEPES 7.5	No	4, 6, 8, 10 & 12	Yes	**20% PEG 4K**
PEG 4K 20% Isopropanol 5 – 20%					**12 % Isopropanol**

The initial conditions that resulted in crystalline mutase were refined to 20% PEG 4K, 12% isopropanol, 0.1M HEPES pH 7.6. These crystals were still not suitable for data collection and further experimentation with these parameters was required. Alternative buffers and alcohols were tried with a view to improving the quality of crystal obtained. Results are listed in Table 2.12.

Table 2.12: Buffer and alcohol optimisation in the crystallisation of mutase.

Precipitant Concentration (w/v) / (v/v)	**Buffer / pH** 0.1M 7.0 – 8.0	**Additive** 10%	**Protein** Concentration mg ml^{-1}	**Crystals**	**Improved Crystals ?**
PEG 4K 20% Isopropanol 12%	Phosphate	No	8	Yes	**No**
	TRIS				
	Na Cacodylate				
	Imidazole				**No**
	Na Acetate				
	MES				
PEG 4K 20%	HEPES 7.6	Butanol			**No**
		Ethanol			
		Methanol			

Although crystalline material was obtained with various other buffers and alcohols no variations produced better crystals and so optimisation continued.

Small molecule additives and detergents are known to interfere and alter the crystallisation characteristics of proteins (Cudney, 1994). Using the commercially available kits from Hampton Research hanging drops were set up with the optimal precipitant concentrations. Additives were used in a ten-fold dilution from the Hampton Kits. Along with the commercial screens known enzyme inhibitors (R. A. Field, unpublished results) were incubated with the protein at room temperature prior to crystallisation. It was anticipated that by 'locking' the enzyme with inhibitor the enzyme would be more susceptible to crystallisation. These crystals would have the added bonus in that if a structure were elucidated using them then the active site of the enzyme would be readily identifiable. Results are listed in Table 2.13.

Table 2.13: Screening additives and detergents with UDP-galactopyranose mutase

Precipitant **Concentration** **(w/v) / (v/v)**	**0.1M Buffer pH 7.6**	**Additive**	**Protein** **Concentration** **mg ml^{-1}**	**Improved Crystals**
20% PEG 4K 12% Isopropanol	HEPES	Hampton A. S. I	8	**Yes, Solution #2**
		Hampton A. S. II		No
		Detergent Screen		No
		2-F-Gal-1-P		No
		1-Deoxygal$_f$		No
		Gal-α-o-methyl		No

Crystals were reproduced with various additives but only addition of L-cysteine at 0.01M yielded crystals of better quality than those already available. However, there were many nucleation sites and crystals were very small, 0.04 x 0.02 x ~0.002mm, though singular with growth appearing to originate from a small dark nucleation site. To improve crystal quality further and increase crystal size, it was evident that further optimisation was required, Table 2.14. Trials with other reducing agents and screening of previously used commercial screens were performed both in the presence and absence of L-cysteine.

Table 2.14: Optimisation of reducing agent present in crystallisation.

Precipitant Concentration (w/v) / (v/v)	**0.1M Buffer pH 7.6**	**Additive 0.001 – 0.1M**	**Protein** Concentration **mg ml^{-1}**	**Mixing**	**Improved Crystals**
PEG 4K 20% Isopropanol 12%	HEPES	L-cysteine	8	Yes	No
		Glutathione			
		DTT			
		β-Mercapto-ethanol			
Hampton I		L-cysteine 0.01M	4, 8, &10	Yes & No	No
Hampton II					
JMac					

Again, although crystallisation was reproducible with both DTT and β-Mercaptoethanol these yielded smaller crystals. Re-screening all the commercial kits in the presence of L-cysteine did not identify any new crystallisation conditions, though many conditions were discovered for the successful crystallisation of L-cysteine.

With optimisation avenues appearing to be at an end, and the crystals not at a size ideal for data collection, alternative methods of crystallisation using the optimised mother liquor as precipitant were explored.

Free Interface Diffusion

In an effort to reduce the number of nucleation sites in crystallisation experiments thereby hopefully growing fewer but larger crystals, a 'novel' technique referred to as free interface diffusion was tried (Hampton Research, 1995). This differs from the more conventional methods of crystallisation in that initially a high local concentration of protein and precipitant exists yielding many nucleation sites, followed by a rapid relaxation to more dilute concentrations suited for crystal growth from pre-existing nuclei (Salemme, 1985).

A thin walled quartz glass capillary of diameter 0.2mm was filled to slightly less than half volume before an equal volume of protein (containing L-cysteine) was pipetted on top of the mother liquor, being careful not to induce any mixing between the two layers. The capillary was sealed and incubated at 20.5°C horizontally. Over several weeks the protein and mother liquor slowly diffused together and yellow crystals of a relatively large (0.5 x 0.01 x 0.01mm) feather like nature were observed. Unfortunately the crystals were of the same morphology as obtained prior to inclusion of L-cysteine and hence were of no use for data collection.

To reduce the rate of diffusion further a small layer of oil was placed between the mother liquor and protein before sealing the capillary. Three different oils (available from Hampton Research), which alter the rate of equilibration differently according to their composition were examined. Al's, paraffin and silicon oil were pipetted on top of the mother liquor in the capillary in various amounts before addition of protein. Although the rate of crystal growth altered, (paraffin, Al's, silicon, diffusion becoming quicker) again no better crystals were observed in relation to the hanging drop experiment.

MicroBatch Crystallisation Under Oil.

Crystallising macromolecules under oil allows one to control the rate of water diffusion through oil and hence the rate at which crystallisation occurs. If paraffin oil is used then no water diffusion is observed and all the reagents involved in the crystallisation remain at specific concentrations (Chayen *et al*, 1992); subsequently no concentration of the protein or the reagents can occur in the drop. However, by using silicon oil one can perform a microbatch experiment that allows for concentration of sample and reagents in the drop as diffusion of water is possible through the silicon oil (D'Arcy *et al*, 1996).

6µl of 100% paraffin oil, or 6µl of a paraffin/silicon oil mix, was pipetted into the well of a 72 well MicroBatch plate (Hampton Research). 2µl of 8 $mgml^{-1}$ protein was pipetted into the cone shaped depression in the MicroBatch plate followed by addition of an equal volume of mother liquor. Trials were set up with different ratios of paraffin to silicon oil to alter the rate of vapour diffusion obtained. The tray was covered to prevent dust contamination and incubated at 20.5°C. After 52 weeks and periodical checking no crystals were observed. This avenue of crystallisation was abandoned.

Seeding

A seed from a crystal too small to be of any use can be used as a template or a nucleation site onto which other 'protein molecules' can add themselves. Larger crystals with the same characteristics as the crystal from which the seed originated should then grow (E. Stura, personal communication). Two seeding techniques, streakseeding and macroseeding, were employed in combination with the hanging drop vapour diffusion method for all trials, (Stura and Wilson, 1992). Using a human hair to touch a crystal grown from the previously

established conditions, a drop containing 2µl protein, 0.5µl L-cysteine and 2.5µl mother liquor was streaked with the hair after it had been used to dislodge seeds from a crystal. Drops that had been prepared immediately prior to streaking did not yield any crystals but if drops were left to equilibrate for approximately 40 hours before streaking then a shower of single crystals resulted within 24 hours of streaking. The seeded crystals were of similar dimensions to those previously obtained but were now free from the polycrystalline nucleation sites. The crystals were still too small to be of use for structure determination but were used in macroseeding experiments.

Single crystals were washed in stabilising solutions of mother liquor 3 to 4 times before being introduced to freshly prepared protein drops, 2µl of 4mg ml^{-1} protein, 0.5µl of 0.01M L-cysteine and 2.5µl mother liquor. Each transfer of crystal was performed using a rayon loop of diameter just larger than the largest dimension of the crystal. Allowing drops to equilibrate for any period of time before introducing crystals did not promote better growth. Typically, crystal growth would be complete after 2-5 days and the macroseeding process then repeated. To obtain crystals of sufficient size for data collection (0.5 x 0.3 x ~0.06 mm), Figure 2.10, seeding had to be repeated 3 to 4 times.

Using this protocol, protein solution was incubated with 2mM FAD both in the absence and presence of NAD(H)/NADP(H) prior to seeding to assess if larger crystals grew as a consequence of the protein being incubated along with co-factor, Table 2.15.

Table 2.15: Incubation conditions and results from subsequent seeding experiments.

Precipitant Concentration (w/v) / (v/v)	**0.1 M Buffer pH 7.6**	**2 mM FAD + 2 mM**	**Incubation** Temperatures °C		**Protein** Concentration mg ml^{-1}	**Improved Crystals ?**	
PEG 4K 20% Isopropanol 12% L-cysteine 0.01M	HEPES	NAD	4	37	4, 8, 12	No	Yes
		NADH	4	37		No	Yes
		NADP	4	37		No	Best
		NADPH	4	37		No	Yes

Crystals were routinely reproduced by this protocol using protein incubated prior to crystallisation at 37°C with FAD and NADP.

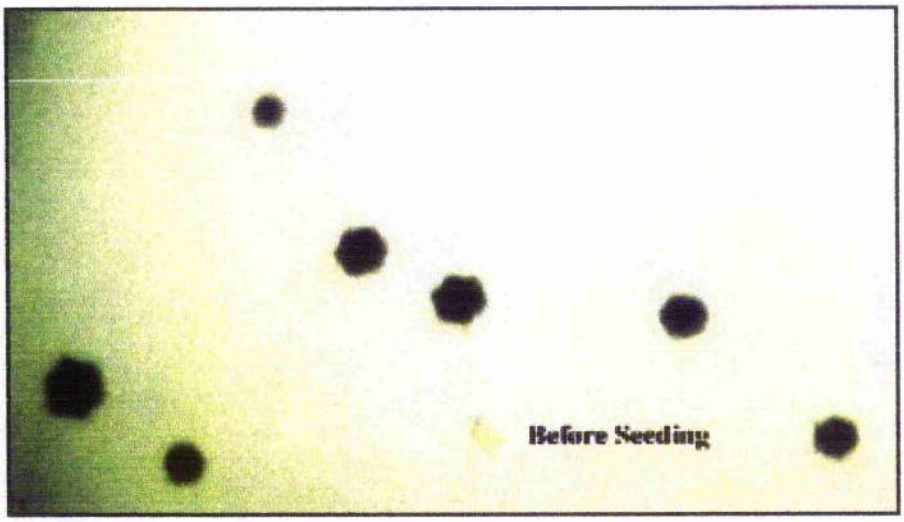
Crystallisation of Selenomethionine Variant Mutase.

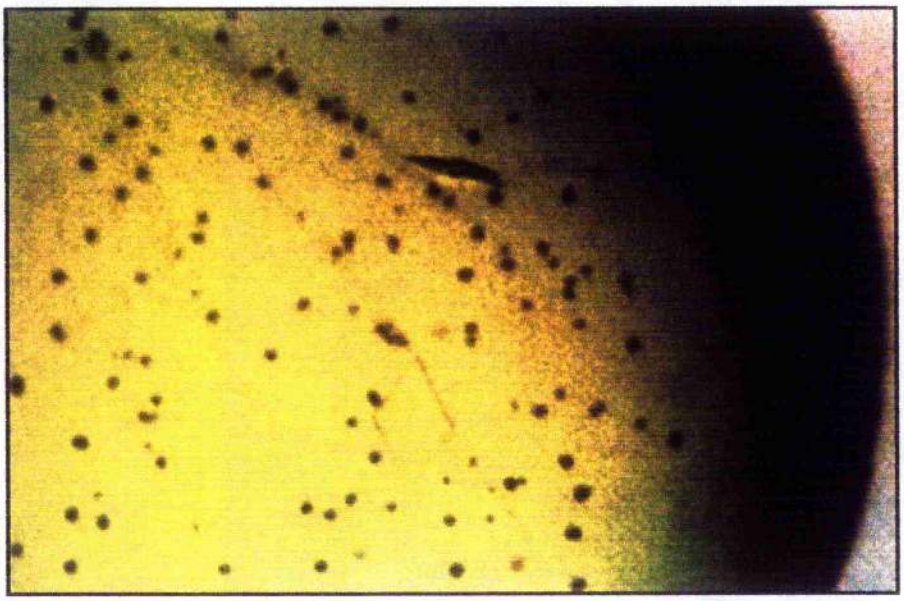
Selenomethionine protein crystallised using the same precipitating reagents and techniques used to crystallise the native enzyme. To ensure the likelihood of oxidation was kept to a minimum all buffers for crystallisation were thoroughly degassed before use. Furthermore, an argon atmosphere was established by attaching an inverted funnel through which argon gas was flowing. All crystallisation experiments involving selenomethionine protein were performed in this atmosphere.

Overall, crystal growth with selenomethionine protein was slower than the native; generally requiring an extra seeding step to get crystals suitable for X-ray diffraction. Even the crystal

size was smaller than the typical size obtained with native protein. The maximum sizes of selenomethionine crystals was 0.3 x 0.15 x ~0.03mm.

Figure 2.10: Crystals of UDP-galactopyranose mutase before and after seeding. The dimensions before seeding are 0.08 x 0.04 x ~0.004mm and increase to 0.5 x 0.3 x ~0.06mm.





Determination of Cryo-protectant Conditions for the Precipitant used to Crystallise Mutase.

Data collection at room temperature is often hampered by the generation of free radicals within the crystal, which limits if not destroys the diffraction obtainable from the crystal. This was observed with the first mutase crystal grown which no longer diffracted after 6 degrees of data collection in-house at room temperature. Data collection at cryogenic temperatures (~ 100K) serves to reduce radiation damage suffered by macromolecular crystals. There are other benefits to data collection at low temperature including a reduction in background scatter, an effective resolution increase, reduction in thermal parameters and the capability to collect all data from one crystal before storage at cryogenic temperature for further use at a later stage (Rodgers, 1997). For data collection with very high intensity beams found at synchrotrons, it is almost essential that macromolecular crystals are suitably cryoprotected.

Good cryoprotectants will neither crack nor increase the mosaic spread of the crystal and will form a 'rigid glass' around the crystal on exposure to cryogenic temperatures. Ice formation should be avoided, as this often degrades both the crystal and the data.

To the mother liquor used to grow crystals, glycerol was added in 2% increments until complete cryoprotection was obtained. This was judged by scooping up cryo-liquor with a rayon loop and positioning it in the centre of the cryostream (Oxford Cryosystems) at 110K. If the cryo-liquor remains transparent then no ice has formed and the solution is suitable for cryoprotection. The integrity of any crystals in the cryoprotectant can only be assessed after exposure to X-rays and comparing the data to room temperature data. Ideally the two pieces of data should scale isomorphously. The minimum percentage of glycerol required for

mutase precipitant was 12% but on occasion ice was observed at this percentage of glycerol. To guarantee complete cryo-protection all the time the concentration of glycerol was increased to15%. MPD and PEG 400 were also assessed as potential cryoprotectants and again 15% was the optimum percent of reagent required.

Native Data Collection.

Two crystal forms were obtained. The first crystal form, grown over one week from a solution of 10mg ml^{-1} protein using 10%PEG 6K (w/w), 0.1M HEPES pH 7.6 as precipitant, dimensions 0.4 x 0.2 x 0.1mm was mounted in a thin-walled glass capillary and sealed in the presence of mother liquor. All diffraction data were recorded at room temperature using the Enraf-Nonius/MacScience DIP2000 dual image plate. X-rays were generated (λ=1.54Å) using an Enraf-Nonius FR591 rotating anode generator and focussed using the MacScience mirror system. An oscillation range of 1° was set, exposure time 30 minutes and the crystal to detector distance set at 150mm. The fresh crystal diffracted to 2.7Å. Radiation damage limited the number of degrees of data collected to 6. The data was indexed using DENZO (Otwinowski & Minor, 1996) in an orthorhombic space group (P222, $P2_12_12_1$, $P2_12_12$ or $P222_1$) with cell dimensions a=56Å b=94Å and c=134Å $\alpha=\beta=\gamma=90°$. As mentioned previously it has not been possible to reproduce this crystal form.

Crystals grown from the seeding procedure have been readily reproduced. Crystals are again yellow, diamond shaped flat plates of average dimensions 0.4 x 0.2 x ~0.05mm. Using the in-house Nonius rotating anode and image plate system as before a crystal of dimensions 0.6 x 0.3 x ~0.05mm was equilibrated in a solution of mother liquor containing 15% glycerol and rapidly transferred to the cold nitrogen stream (110K). An oscillation range of 1° was

employed, exposure time 50 minutes and crystal to detector distance 200mm. The crystal diffracted very weekly to 3.5Å throughout the 180° degrees of data collected. Data were processed using DENZO and SCALEPACK (Otwinowski & Minor, 1996) this time in a monoclinic space group P2 or $P2_1$with unit cell dimension a = 69.544Å, b = 57.126Å, c = 97.555Å, $\alpha = \gamma = 90$, $\beta = 96.44°$. Analysis of systematic absences confirmed the space group as $P2_1$. The data are 98.4% complete from 26Å to 3.5Å, R_{merge} 18.6% and redundancy 1.7. The high R_{merge} is a function of weak data and low resolution. Data collection and reduction statistics are shown in Table 2.16.

The mosaic spread was fixed to 0.5° for scaling, as it proved unstable in refinement. It should be noted that although this crystal diffracted to 3.5Å not all crystals of this size did. The diffracting power of these crystals varied from batch to batch. Some crystals only diffracted to very low resolution ~10Å whilst others would diffract to nearer 3.5Å in-house. As the data was weak an accurate estimation of the mosaic spread was impossible to ascertain.

Table 2.16: Quality of the in-house native data set of UDP-galactopyranose mutase.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Average Redundancy
26.00-7.49	965	92.2	8.8	1.6
7.49-5.97	1014	98.7	13.2	1.6
5.97-5.22	979	99.4	15.4	1.6
5.22-4.74	998	99.2	14.7	1.6
4.74-4.41	994	99.3	15.7	1.7
4.41-4.15	988	99.1	18.4	1.7
4.15-3.94	983	99.2	22.2	1.7
3.94-3.77	981	99.0	27.0	1.7
3.77-3.62	984	99.4	31.0	1.7
3.62-3.50	989	99.3	36.9	1.7
25.00-3.50	**9875**	**98.4**	**18.6**	**1.7**

A higher resolution data set was collected on BM14 (λ=1.001Å) using the MAR345, image plate at the European Synchrotron Radiation Facility (E.S.R.F.), Grenoble, France. The MAR was set in the 18cm mode and the slits were set to 200µm x 200µm. Data collection proceeded as 145 non-overlapping images in steps of 1° at a crystal to detector distance of 242mm. Each exposure was for 120 seconds. For images 1 – 25 the beam current was 138.9mA, ion chamber 7.7. The beam was refilled after image 25. Thereafter the current slowly decayed from 200mA ion chamber 11.3 to a current of 185.6mA and ion chamber 10.4 at the end of data collection. Diffraction was visible to 2.8Å, though weak below 2.9Å, Figures 2.11 and 2.12. No significant crystal decay was observed. Data were processed using DENZO and SCALEPACK (Otwinowski & Minor, 1996) yielding 17660 unique reflections. The crystal was indexed in a monoclinic space group with unit cell dimensions a

= 71.12Å, b = 58.42Å, c = 97.89Å, $\alpha = \gamma = 90$, $\beta = 96.38$. Analysis of diffraction data identified systematic absences consistent with space group $P2_1$. The mosaic spread refined to 2.3° in SCALEPACK. Assuming a dimer of molecular mass 43kDa per monomer, the V_m is $2.35Da^{-3}$ which falls in the allowed range for protein crystals (Matthews, 1968) and indicates a solvent content of 47%. The data are 92% complete from 15Å to 2.9Å, with an R_{merge} 5.0% and redundancy 2.5. The average I over σ for the data is 16.6. Data collection and reduction statistics are shown in Table 2.17.

Figure 2.11: Section of a 1° oscillation diffraction pattern from a crystal of UDP-galactopyranose mutase taken at BM14, E.S.R.F., Grenoble recorded with a MAR 345 image plate. The resolution at the edge of the photograph corresponds to 2.8Å.

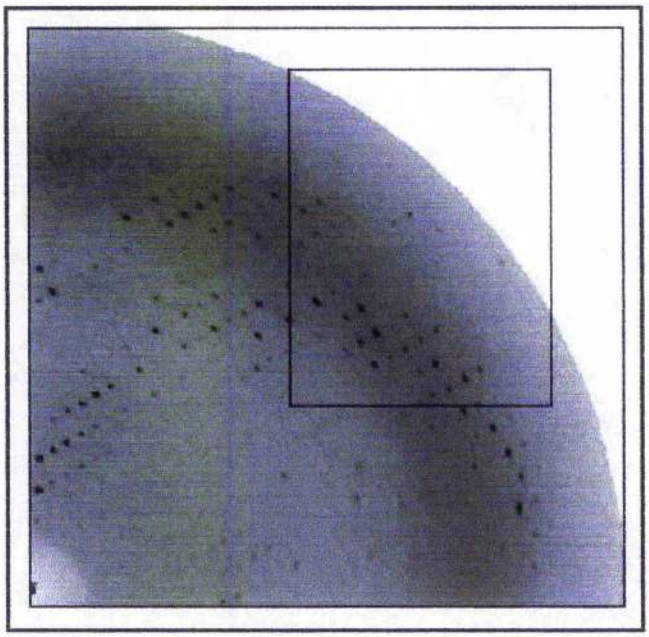


Figure 2.12: High resolution section of a 1° oscillation diffraction pattern of native mutase. Diffraction is visible to 2.8Å at the edge but weak beyond 3.0Å.

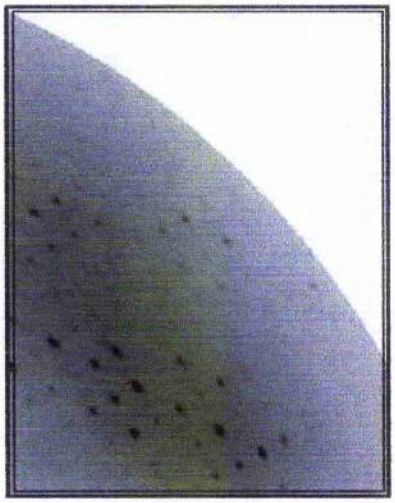


Table 2.17: Quality of the native data set of mutase collected on BM14 using the MAR345, at the E.S.R.F., Grenoble, France.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
15-6.12	1489	82.0	2.6	2.89
6.12-4.91	1678	95.2	2.6	2.89
4.91-4.31	1676	94.5	3.1	2.89
4.31-3.92	1649	95.3	3.8	2.89
3.92-3.65	1654	95.7	4.7	2.89
3.65-3.43	1674	94.3	6.6	2.89
3.43-3.26	1562	90.9	11.4	2.89
3.26-3.12	1556	88.3	19.7	2.89
3.12-3.00	1495	86.0	24.0	2.89
3.00-2.90	1365	80.0	27.7	2.88
15-2.90	**15798**	**90.2**	**5.0**	**2.89**

Screening for Heavy Atom Derivatives.

As no homologous structure is known, phases have to be determined experimentally. To this end a search for isomorphous derivatives was undertaken in-house. The effectiveness of this search was limited by the variability of crystal quality from batch to batch and the effective resolution to which even the best crystals diffracted to in-house (~3.5Å). If encouraging results were obtained (as judged from data reduction and scaling with DENZO and SCALEPACK) in-house then the crystals were stored in liquid nitrogen for data collection at a later date using a synchrotron source. A summary of results obtained in-house is shown in Table 2.18.

Assessment of any potential derivative was primarily carried out in SCALEPACK as an indication of the usefulness of any crystal could be obtained after collection of only a few

frames of data. Examination of the χ^2 and R_{merge} output in the last page of the SCALEPACK log file after scaling a few frames of 'derivative' data against a native data set gives an indication of how good a derivative may be. Crystals that scale closely with the native, that is $\chi^2 \approx 1$, will not be useful derivatives, as there is little statistical difference between the data sets. If $\chi^2 \approx 10$ and $R^2 \approx R$, then this indicates a potential derivative as the differences measured are much larger than the expected error. If χ^2 is much larger than 10 then non-isomorphism may be a problem, (Gewirth, 1996). Results that looked promising in SCALEPACK were later assessed in SCALEIT (CCP4, 1994).

Table 2.18: Searching for isomorphous heavy atom derivatives.

Salt	Concentration (mM)	Length of Soak (Hrs)	χ^2	R %	R^2 %	Comments
K_2PtCl_4	10	0.1				Crystal Dissolved
K_2PtCl_4	5	1				No Diffraction
K_2PtCl_4	2.5	6	3	19	21	2 min b/s[a]
K_2PtCl_4	4	6	4	21	23	2 min b/s
K_2PtCl_4	4	18	4	20	24	2 min b/s
K_2PtCl_4	4	36	3	21	22	2 min b/s
K_2PtCl_6	10	0.1	---	---	---	Crystal Dissolved
K_2PtCl_6	5	6	---	---	---	No Diffraction
K_2PtCl_6	2.5	6	2	19	21	2 min b/s
K_2PtCl_6	3.3	6	4	16	16	2 min b/s
K_2PtCl_6	3.3	20	7.5	23	25	2 min b/s
K_2PtCl_6	3.5	13	3	13	15	4 min b/s
K_2PtCl_6	3.6	20	8	20	21	2 min b/s
K_2PtCl_6	3.6	13.5	5.4	18	21	2 min b/s
K_2PtCl_6	3.6	24	4	19	20	3 min b/s
K_2PtCl_6	3.7	12.5	3	20	24	3 min b/s
K_2PtCl_6[b]	3.8	9	12	14	18	2 min b/s
K_2PtCl_6	3.9	11	11	16	18	2 min b/s
$PtCl_2(NH_3)_2$	4.0	13	3.8	18	21	2 min b/s
$PtCl_2(NH_3)_2$	4.5	18.5	4.5	19	20	2 min b/s
H_2PtCl_2	4.0	13	3	13	16	4 min b/s

Table 2.18(cont): Searching for isomorphous heavy atom derivatives.

Salt	Concentration (mM)	Length of Soak (Hrs)	χ^2	R %	R^2 %	Comments
Hg OAc	5	6	---	---	---	Crystal Cracked No Diffraction
Hg OAc	2.5	1	---	---	---	Crystal Cracked No Diffraction
Hg OAc	0.5	12	---	---	---	Mosaic Pattern
Hg OAc	0.25	1	---	---	---	Mosaic Pattern
Hg OAc	0.05	4	5	24	26	2 min b/s
Hg OAc	0.025	12	2	21	23	2 min b/s
Thimerosal[c]	5	0.1				Crystal Cracked No Diffraction
Thimerosal	2.5	1				Crystal Cracked No Diffraction
Thimerosal	0.5	6				Mosaic Pattern
Thimerosal	0.1	3	3	20	25	2 min b/s
Thimerosal	0.05	6	4	19	20	2 min b/s
Thimerosal	0.05	24	5	24	26	2 min b/s
Ur OAc	2.5	1	---	---	---	Crystal Dissolved
Ur OAc	0.06	18	4	18	20	2 min b/s
Ur OAc	0.02	18	4	21	23	2 min b/s

[a] Backsoak
[b] Collected at the E.S.R.F.
[c] Ethyl (2-mercaptobenzoate-S) mercury sodium salt

Data Collection with Crystals Soaked in K_2PtCl_6.

A putative platinum derivative was identified after soaking a crystal for 9.5 hours in 3.8mM K_2PtCl_6. The crystal was back soaked in cryo-protectant for 2 minutes before being rapidly transferred to the cold nitrogen stream. 19° of data were collected at the E.S.R.F. to 3.0Å and processed and scaled using DENZO and SCALEPACK. The data refined to a $\chi^2 = 11.4$ against the native data set previously collected at the E.S.R.F., Table 2.19, which is indicative of a potential derivative (Gewirth, 1996). Time limitations at the synchrotron prevented the collecting of a more complete data set.

Table 2.19: Analysis of χ^2 and R_{merge} from SCALEPACK after scaling three frames of data collected from a crystal soaked in 3.8mM K_2PtCl_6 against a native data set.

Å	χ^2	R	R^2
25.00 - 8.31	29.5	0.164	0.244
8.31 - 6.64	12.8	0.128	0.183
6.64 - 5.81	16.4	0.151	0.191
5.81 - 5.28	10.9	0.135	0.165
5.28 - 4.91	10.9	0.135	0.173
4.91 - 4.62	10.8	0.134	0.173
4.62 - 4.39	1.8	0.136	0.199
4.39 - 4.20	6.8	0.123	0.159
4.20 - 4.04	7.3	0.165	0.216
4.04 - 3.90	7.6	0.201	0.238
All Reflections	**11.4**	**14.0**	**18.5**

The above experiment was repeated on a return journey to the E.S.R.F. and this time a full data set recorded. Data were collected as 298 non-overlapping images at 0.5° oscillations on

station ID14 EH3 using a MAR CCD with a crystal to detector distance of 148mm and λ = 0.926Å. Each exposure lasted 120 seconds. The data were reduced and scaled using MOSFLM (Leslie, 1993) and SCALA (Evans, 1993). Scaling of 10° of the data to native data in SCALEPACK for comparison purposes revealed a $\chi^2 = 12$, R = 8.3% and $R^2 = 9.1\%$. The mosaic spread of the crystal refined to 2.0°. The data are 81.4% complete to 3.5Å, R_{merge} 9.1% and redundancy 2.8, Table 2.20.

Using the programs SORTMTZ, TRUNCATE, CAD and SCALEIT as implemented in CCP4 (CCP4, 1994), the data (20 – 3.5Å) were converted into a format suitable for difference Patterson map calculation. Isomorphous difference maps were calculated and plotted using NPO (CCP4, 1994), and viewed at section y = ½, Figure 2.13. Patterson sites were calculated by hand and refined in MLPHARE, (CCP4, 1994). None of the sites refined any better than a random x, y, z, coordinate and subsequently the positions of the heavy atoms could not be deduced from this data.

Figure 2.13: Isomorphous difference Patterson map, Pt derivative – native. All data collected at the E.S.R.F., Grenoble, France.

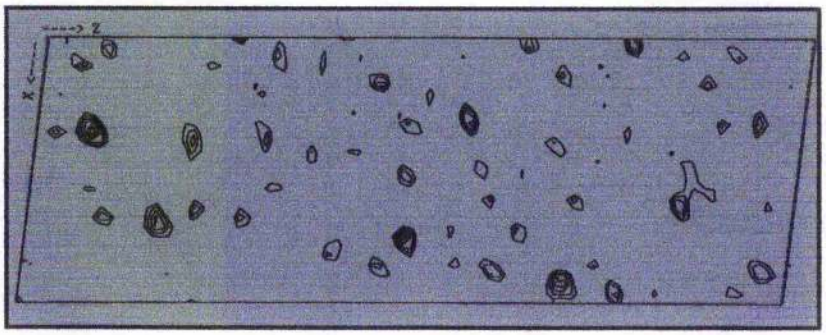


Table 2.20: Quality of the platinum data set collected at ID14, EH3 using a MARCCD, at the E.S.R.F., Grenoble, France.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Average Redundancy
15-9.49	530	70.8	7.3	2.3
9.49-6.71	1300	77.8	6.1	2.8
6.71-5.48	1747	80.3	6.1	2.9
5.48-4.74	2136	82.4	7.3	2.9
4.74-4.24	2314	82.8	6.7	2.8
4.24-3.87	2630	83.7	6.9	2.8
3.87-3.59	2791	84.0	8.3	2.8
3.59-3.35	3017	84.1	9.9	2.9
3.35-3.26	3294	84.7	12.7	2.8
3.16-3.00	3345	85.2	12.7	2.9
15-3.00	**23104**	**81.4**	**9.1**	**2.8**

Two platinum salts, K_2PtCl_4 (Pt^{II}) and K_2PtCl_6 (Pt^{IV}), were tested for their potential to bind isomorphously to mutase crystals. Early experiments suggested K_2PtCl_6 was binding with somewhat more affinity judging from the SCALEPACK output. Binding parameters were refined to 3.8mM K_2PtCl_6 for 9.5 hours, though soaking crystals in concentrations between 3.5 and 4.0mM largely gave the same results when soaked for the same length of time. Promising as the early indications were, we were not able to locate heavy atom positions with data from these 'derivatives'. Instead of a limited number of platinums binding specifically, many appear to be binding, contributing to the noisy Patterson maps obtained. The high mosaic spread is also a compounding factor.

As an alternative to platinum, two mercury salts were assayed for their potential to form isomorphous derivatives. Unlike platinum, mercury is known to bind specifically to cysteine

residues (mutase contains two cysteines) so a problem of low occupancy is less likely. Initial experiments that never destroyed the crystals when judged by eye gave highly mosaic diffraction patterns on exposure to X-rays. However, lowering the working concentration removed this problem and indeed yielded some promising results. Results from SCALEPACK suggest binding is taking place; that the crystals crack and become mosaic at higher concentrations support this notion. More experiments are required to optimise the soaking conditions. Likewise, soaking experiments with uranium acetate have yielded similar results. Unlike mercury, uranium does not bind to any specific amino acid, hence an outcome similar to the platinum experiments is more likely.

Data collection with crystals 'soaked' in Xenon

If protein crystals are pressurised with xenon gas under the appropriate conditions it is possible to observe xenon binding to the protein (Sauer, 1997). Xenon tends to bind weakly in hydrophobic pockets reducing the problem of non-isomorphism, a problem often encountered when heavy metals are used. Several derivatives can potentially be produced from one crystal as xenon often binds reversibly, allowing the same crystal to be used in heavy metal soaking experiments. Similarly, as the pressure at which binding is performed controls the number of xenon atoms that bind, multiple xenon derivatives can be had from the same crystal.

A crystal of dimensions 0.4 x 0.25 x ~0.005mm was soaked in cryoprotectant and scooped up inside an appropriately sized rayon loop. The loop was attached to magnetic head and attached to the lid of the pressurisation chamber and sealed in a closed system. 200µl of mother liquor was present within the chamber to prevent the crystal from drying out. The

system was flushed with xenon gas before pressurising to 80 psi. This pressure was maintained for 20 minutes before a controlled release of xenon pressure, and the crystal quickly mounted in the cryostream at 110K. The apparatus used in this experiment was designed and built in-house. A schematic representation of the apparatus is shown in Figure 2.14 along with a cross section of the pressurisation chamber in Figure 2.15.

5° of data were recorded at 110K using the Enraf-Nonius/MacScience DIP2000 dual image plate in-house. An oscillation range of 1° was fixed, exposure time 50 minutes and the crystal to detector distance set at 200mm. The crystal diffracted weekly to 3.7Å and the data were processed in DENZO and SCALEPACK. The χ^2 refined to 5 and the linear R_{merge} 26%, square R_{merge} 33%.

Encouraging though these early results were, it should be noted they were not achieved at the first attempt. Crystals often showed no diffraction and some diffracted more weakly than normal. It was felt that more reproducible results were guaranteed by other techniques, namely heavy metal soaking and selenomethionine substitution.

Figure 2.14: Aerial view of a schematic representation of the xenon pressure cell and components.

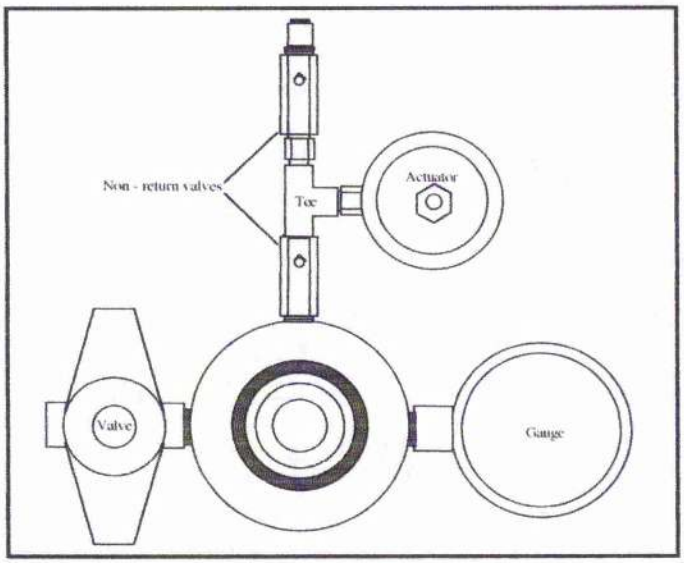
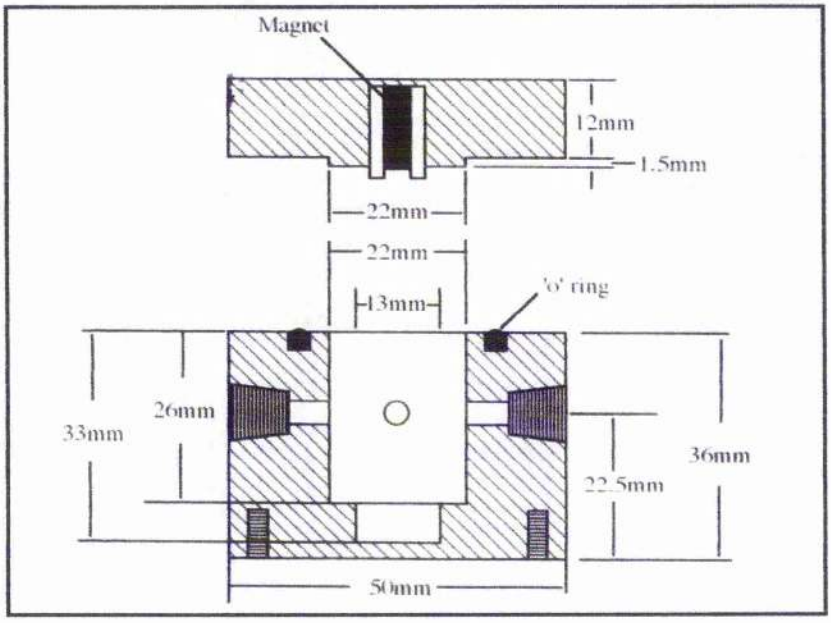


Figure 2.15: Cross section of the xenon pressurisation chamber and lid.



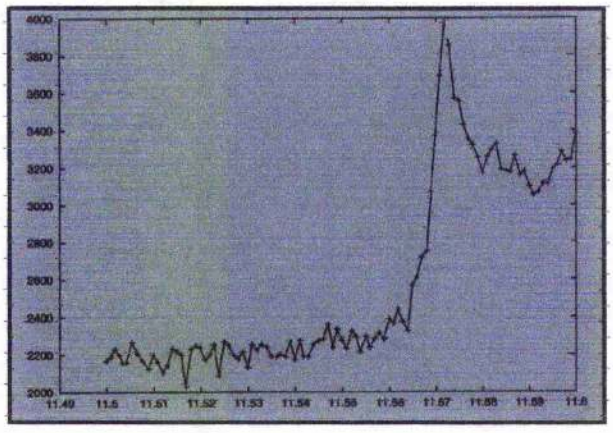
MAD at BM14

Multiwavelength Anomalous Dispersion (MAD) was examined as opposed to MIR as an alternative route to structure solution. Experiments were conducted at the E.S.R.F. with crystals containing either the anomalous scattering atom platinum or selenium.

A crystal of dimensions 0.5 x 0.25 x ~0.05mm was soaked in 3.8mM K_2PtCl_6 for 10 hours and back soaked in cryo-protectant before being transferred to the cold nitrogen stream. An EXAFS scan detected the presence of platinum within the crystal, Figure 2.16.

The first data set (pte1) was collected on the platinum white line at λ = 1.0714 Å (E = 11.57KeV). Data were collected as 150 non-overlapping images using a MAR345 detector in 18cm mode, in steps of 1° at a crystal to detector distance of 220mm. Data were processed using DENZO and SCALEPACK to 3.0 Å (Otwinowski & Minor, 1996) yielding 15202 unique reflections. The crystal was indexed in a monoclinic space group with unit cell dimensions a = 68.84 Å, b = 56.35 Å, c = 97.42 Å, $\alpha = \gamma = 90°$, β = 96.62°. The data are 90.4% complete and the R_{merge} is 6.1%. Data quality for individual resolution bins is shown in Table 2.21a.

Figure 2.16: EXAFS scan of crystal soaked in K_2PtCl_6 for 10 hours and used in data collection.



A second data set (pte2) was collected at the point of inflexion of the platinum edge λ = 1.0714Å (E = 11.57KeV). All other data collection parameters were consistent with pte1 collection. Data were again processed using DENZO and SCALEPACK to 3.0Å (Otwinowski & Minor, 1996) yielding 15202 unique reflections. The crystal was indexed in a monoclinic space group with unit cell dimensions a = 68.88Å, b = 56.41Å, c = 97.44Å, α = γ =90°, β = 96.60°. The data are 94.0% complete and the R_{merge} is 5.3%. Data quality for individual resolution bins is shown in Table 2.21b.

The third data set (pte3) was collected on the high energy side of the platinum edge, λ = 0.8855Å (E = 14.00KeV). All other data collection parameters were the same as pte1 collection. Data were again processed using DENZO and SCALEPACK to 3.0Å (Otwinowski & Minor, 1996) yielding 15265 unique reflections. The crystal was indexed in a monoclinic space group with unit cell dimensions a = 68.94Å, b = 56.46Å, c = 97.56Å, α = γ =90°, β = 96.60°. The data are 92.8% complete and the R_{merge} 5.6%, Table 2.21c. The mosaic spread refined to 1.7° in SCALEPACK during scaling of each data set.

Tables 2.21: Quality of the MAD data sets denoted pte1, 2 & 3. Collected at the E S R F, Grenoble, France.

a) Collected on the platinum edge

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
30.00-7.36	1577	80.8	2.3	1.6
7.36-5.86	1713	87.3	3.9	1.6
5.86-5.12	1697	89.0	4.2	1.6
5.12-4.66	1753	89.1	4.0	1.6
4.66-4.32	1735	89.8	3.9	1.6
4.32-4.07	1776	90.8	4.4	1.6
4.07-3.87	1800	90.8	5.3	1.6
3.87-3.70	1745	91.3	6.6	1.6
3.70-3.56	1820	92.5	8.1	1.6
3.56-3.43	1828	92.7	9.8	1.6
3.43-3.33	1760	92.8	13.1	1.6
3.33-3.23	1839	93.3	16.2	1.6
3.23-3.15	1780	93.5	20.0	1.6
3.15-3.07	1804	91.3	23.3	1.6
3.07-3.00	1782	91.4	27.2	1.6
30.00-3.00	**26409**	**90.4**	**6.1**	**1.6**

b) Collected on the point of inflexion

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
30.00-7.36	1634	83.7	2.1	1.6
7.36-5.86	1812	92.4	3.3	1.6
5.86-5.12	1791	94.0	3.7	1.6
5.12-4.66	1849	94.0	3.2	1.6
4.66-4.32	1819	94.2	3.4	1.6
4.32-4.07	1860	95.1	4.0	1.6
4.07-3.87	1883	95.0	4.7	1.6
3.87-3.70	1819	95.1	6.0	1.6
3.70-3.56	1887	95.9	6.8	1.6
3.56-3.43	1880	95.4	8.0	1.6
3.43-3.33	1825	96.2	10.8	1.6
3.33-3.23	1885	95.6	13.2	1.6
3.23-3.15	1831	96.2	27.8	1.6
3.15-3.07	1845	93.4	21.3	1.5
3.07-3.00	1831	93.9	23.1	1.5
30.00-3.00	**27451**	**94.0**	**5.3**	**1.6**

c) Collected on the high energy side of the platinum edge.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
30.00-7.36	1608	82.4	2.3	1.5
7.36-5.86	1774	89.9	3.3	1.5
5.86-5.12	1770	92.1	3.9	1.5
5.12-4.66	1842	93.0	3.3	1.5
4.66-4.32	1801	93.4	3.8	1.5
4.32-4.07	1828	94.0	4.2	1.5
4.07-3.87	1901	93.6	5.1	1.5
3.87-3.70	1789	93.7	6.3	1.5
3.70-3.56	1888	95.6	7.5	1.6
3.56-3.43	1853	93.7	9.1	1.6
3.43-3.33	1799	95.6	11.2	1.6
3.33-3.23	1902	95.1	15.3	1.6
3.23-3.15	1810	94.9	18.0	1.6
3.15-3.07	1830	91.7	22.9	1.6
3.07-3.00	1817	93.3	25.1	1.6
30.00-3.00	**27212**	**92.8**	**5.6**	**1.6**

Data Analysis

The scaled data were taken through SCALEPACK2MTZ, CAD and TRUNCATE as implemented in CCP4 (CCP4, 1994) to convert it into a format suitable for difference Patterson map calculation. Both anomalous and dispersive difference maps were calculated and plotted using NPO (CCP4, 1994) and viewed at section y = ½, Figures 2.17 and 2.18. Heavy atom coordinates were deduced using RSPS and refined in both MLPHARE (CCP4, 1994) and SOLVE (Terwilliger and Berendzen, 1999). Occupancy levels refined to 0.4 at best, consequently phases could not be extracted from this data. Rather than having few high occupancy platinum sites it appears there are many sites with low occupancy. This explains the promising results acquired with this data at the processing level from SCALEPACK and the noisy Patterson maps generated.

Figure 2.17: Dispersive difference Patterson map, pte1 – pte2. All data collected at the E.S.R.F., Grenoble, France.

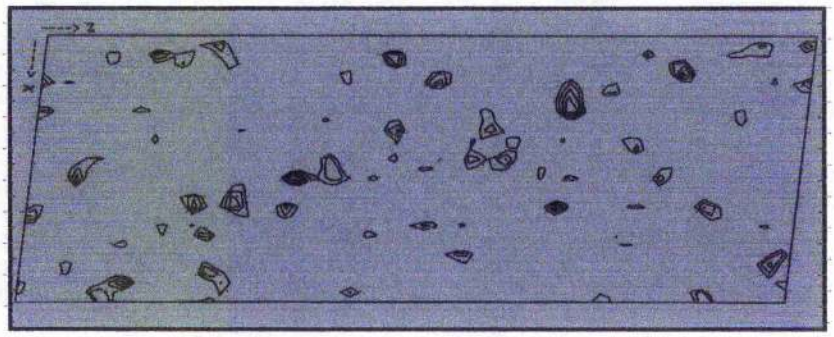
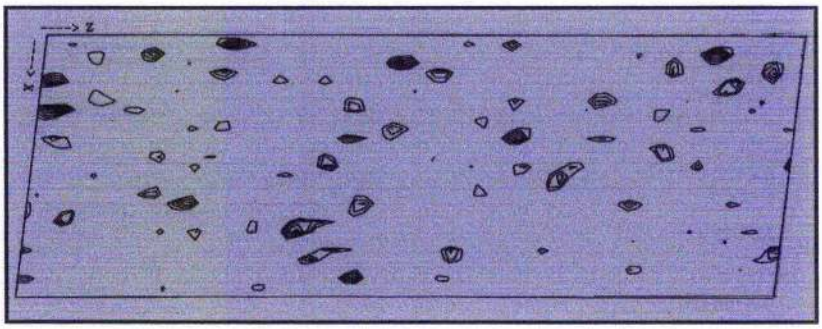


Figure 2.18: Anomalous difference Patterson map, native - pte1. All data collected at the E S R F, Grenoble, France



Data Collection with Selenomethionine Variant Protein Crystals.

An EXAFS scan, Figure 2.19, was performed on a selenomethionine variant protein crystal to confirm selenium presence, and a data set subsequently collected on ID14, EH3, at the E.S.R.F. A single crystal of dimensions 0.4 x 0.25 x ~0.03mm was used to record the entire data set in two sweeps of 25 seconds per 1°. A MARCCD was employed and the crystal to detector distance was set to 148mm. The data was reduced and scaled using MOSFLM (Leslie, 1993) and SCALA (Evans, 1993). The data were 94% complete (74% of the anomalous data, R_{anom} 6.8%), consisting of 12293 unique reflections to 3.0Å yielding R_{merge} 6.8% and redundancy 1.8. The mosaic spread refined to 1.5°. This was encouraging, as at this level the mosaic spread would probably have been manageable. Table 2.22 shows the data collection statistics for individual resolution bins.

Figure 2.19: EXAFS scan of crystal grown in defined media supplemented with selenomethionine and subsequently used in data collection.

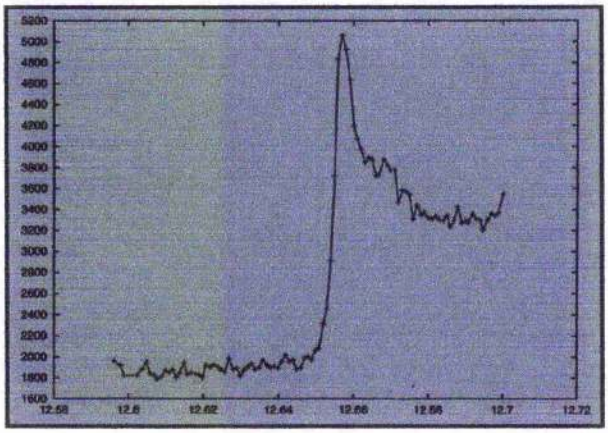


Table 2.22: Quality of the Se-Met data collected on ID14, EH3 using a MARCCD, at the E. S. R. F., Grenoble, France.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Average Redundancy
15-9.49	666	87.9	4.9	1.7
9.49-6.71	1353	91.9	3.9	1.8
6.71-5.48	1745	93.3	6.3	1.8
5.48-4.74	2090	94.5	4.5	1.8
4.74-4.24	2341	94.6	4.7	1.9
4.24-3.87	2570	94.6	8.3	1.9
3.87-3.59	2776	94.5	5.8	1.8
3.59-3.35	2976	94.3	7.2	1.9
3.35-3.26	3229	94.5	13.6	1.9
3.16-3.00	3337	94.4	13.7	1.9
15-3.00	**23084**	**94.4**	**6.8**	**1.8**

Using the programs, SORTMTZ, TRUNCATE, CAD and SCALEIT as implemented in CCP4 (CCP4, 1994) the data (15 – 3.0Å) were converted into a format suitable for Patterson map calculation. The native data set collected to 3.0Å on a previous visit to Grenoble was

used. Both anomalous and isomorphous difference maps were calculated and plotted using NPO (CCP4, 1994) and viewed at section y = ½. Selenium coordinates were deduced using RSPS (CCP4, 1994) and refined in MLPHARE (CCP4, 1994). Again Patterson maps were noisy with the highest observable peak at 4σ. The positions of the seleniums could not be deduced from this data.

A second MAD experiment was attempted at a latter date also at Grenoble when the beam was operating at a reduced power level. Despite mounting in excess of 30 selenomethionine variant protein crystals not one diffracted to sufficient resolution together with a low enough mosaic spread. If smaller crystals were used, 0.1 x 0.05 x ~0.005mm, then although the resolution was low and the diffraction weak, the observed mosaic spread was manageable (<1.7°).

UDP-galactopyranose Mutase from Klebsiella pneumonia

Chris Whitfield, University of Guelph, donated approximately 20mgs of purified mutase from *K. pneumonia*. The protein was shipped in 50% glycerol and displayed an intense yellow similar to mutase from *E. coli*. The protein was dialysed for 3 x 40 minutes against 50mM TRIS, 2mM DTT. pH 7.6 in preparation for dynamic light scattering and crystallisation.

2mL of 50mM TRIS, 2mM DTT, pH 7.6 was 0.02μm filtered and injected through the optical cell of the Dynapro 801 to remove dust and air particles before injecting protein sample. 500μl of *K. pneumonia* mutase was diluted to 6mg ml^{-1}, 0.02μm filtered and injected into the optical cell. The protein count rate was monitored until stabilising around 690. At this point measurement commenced. The results are displayed in Table 2.23.

Table 2.23: Dynamic light scattering results at room temperature for mutase from K. pneumonia. The average polydispersity value as a function of the hydrodynamic radius is 27.7%.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDa	Temp. C	Count Rate	Base Line	PolyD %
1	0.815	351	6.1	1.777	236	20.7	629	1.000	29.1
2	0.814	337	6.3	2.222	253	20.2	673	1.000	35.2
3	0.807	352	6.1	1.718	236	20.7	690	1.001	28.1
4	0.802	352	6.1	1.710	233	20.5	692	1.000	28.0
5	0.793	350	6.0	1.759	231	20.2	694	1.000	29.3
6	0.783	345	6.2	1.733	245	20.7	701	1.000	27.9
7	0.779	343	6.2	1.807	246	20.6	699	1.000	29.1
8	0.776	355	6.0	1.748	230	20.4	695	1.001	29.1
9	0.789	352	6.1	1.390	236	20.6	692	1.002	22.8
10	0.807	349	6.1	1.145	232	20.7	692	0.009	18.7

As the sample was not monodisperse at room temperature this experiment was repeated at 4°C to judge if crystallisation at a lower temperature may be more profitable. The sample was recovered by drawing back on the syringe and used for the experiment the following morning after the Dynapro 801 had been stored at 4°C overnight. As before, 500µl of *K. pneumonia* mutase was filtered and injected into the Dynapro 801. The protein count rate was monitored and on stabilisation measurement began. The results are illustrated in Table 2.24.

Table 2.24: Dynamic light scattering results at 10°C for mutase from K. pneumonia. The average polydispersity value as a function of the hydrodynamic radius is 47.1%.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.M W kDa	Temp. °C	Count Rate	Base Line	PolyD %
1	0.775	223	7.2	3.459	353	10.3	574	1.035	48.0
2	0.785	236	6.7	3.335	300	9.9	541	1.010	49.7
3	0.768	215	7.3	3.882	361	9.4	589	1.026	53.1
4	0.730	194	8.1	4.542	474	9.6	648	1.053	56.1
5	0.792	231	6.8	3.277	309	9.6	552	1.014	48.1
6	0.733	228	6.9	3.398	314	9.4	571	1.018	49.2
7	0.782	243	6.5	2.552	274	9.6	541	1.010	39.2
8	0.790	243	6.4	2.534	270	9.4	533	1.005	39.5
9	0.784	247	6.4	2.505	262	9.6	534	1.005	39.1
10	0.767	242	6.5	3.234	278	9.7	534	1.008	49.7

A simple comparison of the average polydispersity as a function of the hydrodynamic radii reveals that crystallisation would be best attempted at room temperature. Recalling that if crystallisation is likely, then this figure should be lower than 25%. At 27.7% crystallisation is not guaranteed but appears more likely than at 4°C.

Analysis of the protein by SDS-PAGE revealed more than one band on the gel. Without further purification the protein solution was estimated to be around 85% homogenous. As any further purification at this stage would result in loss of valuable protein it was decided to pursue crystallisation with the protein in this condition.

Crystallisation of UDP-galactopyranose Mutase from K. pneumonia

The sitting drop vapour diffusion method was employed throughout all crystallisation trials of *K. pneumonia* mutase. Initially, commercial crystallisation screens from Hampton Research along with the JMac and PEG, PEG/LiCl, MPD, $(NH_4)_2SO_4$ and NaCl grid screens were used to screen for initial crystallisation conditions with protein at 12 and 6mg ml^{-1}. Crystallisation agent number 100 from JMac, 15% PEG 4K, 0.2M $MgCl_2$, 0.1M HEPES pH 7.5 proved successful in crystallising this protein at 6mg ml^{-1}. Crystals grew from precipitate as large clusters of thin needles, dimensions 0.2 x 0.01 x 0.01mm. The crystals were not suitable for X-ray diffraction and optimisation of these conditions was required. To this end the pH was varied from 4 - 9 and similarly the PEG 4K concentration from 2 – 21%, Table 2.25.

Table 2.25: Different buffers and pH values tried against 2 - 21% PEG 4K in optimising the original crystals obtained of mutase from K. pneumonia.

Buffer (0.1M)	PH	Crystals
Citric	4.0	No
MES	4.0 – 6.5	No
Sodium acetate	5.0	No
Sodium aitrate	5.6	No
Imidazole	7.0	Yes
Sodium acetate	7.0	Yes
HEPES	7.0 – 7.5	Yes
MES	7.0	Yes
Sodium cacodylate	7.5	Yes
TRIS	7.5 – 9.5	Yes
Bicine	9.0	Yes

At pH levels < 7.0 the enzyme failed to crystallise under these conditions, however above pH 7 the enzyme readily crystallised with a variety of buffers. The optimum parameters for crystallisation were refined to 20% PEG 4K, 0.2M $MgCl_2$, 0.1M TRIS pH 8.5. In place of the large clusters of needles that were first obtained, single crystals were now evident. However, the yellow crystals are essentially still only two dimensional, the largest dimension being 0.4mm and the other two dimension being in the μm range.

A single crystal of approximately 0.3 x 0.01 x 0.01mm was mounted in a thin walled glass capillary in the presence of the mother liquor and mounted on the in-house Enraf-Nonius/MacScience DIP2000 dual image plate using an Enraf-Nonius FR951 rotating anode generator. The crystal was exposed to X-rays for 50 minutes over 1° oscillation range. No diffraction was observed.

That a crystal of these dimensions does not show any diffraction is not unusual. The crystals are very small. One would prefer to increase the smallest dimensions by a minimum factor of 10 or 50 - 100 ideally. Free radical generation may destroy the diffraction capabilities almost immediately, thus making cryoprotection essential. The crystals grown thus far have been obtained after limited optimisation. However, there are many other parameters that remain to be optimised.

As the protein was not overexpressed and purified in-house it may be more beneficial to overexpress and purify the protein ourselves and try to and achieve a higher state of protein purity before embarking on crystallisation. Analysis by mass spectrometry should be performed where possible, to ensure any sample is of maximum purity before attempting crystallisation. This analysis was omitted in the initial characterisation of this protein.

We have recently received the clone for UDP-galactopyranose mutase from *K. pneumonia* from Chris Whitfield and work on the overexpression and purification of the enzyme in-house is on-going.

2.4 Discussion and Conclusions

UDP-galactopyranose mutase has been overexpressed and purified to suitable homogeneity such that crystallisation of the protein is viable. Characterisation of the protein by electrophoresis and mass spectrometry has shown the protein to be greater than 95% homogenous, although three peaks are evident by mass spectrometry suggesting some heterogeneity. Dynamic light scattering results are consistent with a dimer of molecular mass 86kDa and data from non-denaturing gels confirm this data.

The purified protein displays a yellow colour due to the associated co-factor, FAD. It is also postulated that other co-factors may be present, NAD(H) or NADP(H) and indeed preincubation of the protein with two cofactors (FAD + NAD or NADP) yields better quality crystals. The protein from *K. pneumonia* is known to require NADP for activity.

Two crystal forms of UDP-galactopyranose mutase from *E. coli* have been grown in a form suitable for x-ray diffraction. One crystal form is orthorhombic and diffracted to 2.7Å in-house. This crystal has not been reproduced despite extensive efforts. The inability to reproduce this crystallisation is attributed to multiple isoforms of the protein being present when the protein is overexpressed in LB. In combination with the possibility of co-factors being partially stripped from the protein this makes the crystallisation process more serendipitous than one would desire. To try and overcome these problems several expression media were examined for yield of soluble protein produced. Terrific broth was chosen as the

best suitable media. Protein yield was increased eight fold and the multiple isoforms evident from LB were removed.

An alternative monoclinic crystal form has been reproducibly grown in hanging drops from protein overexpressed in terrific broth. However, to obtain crystals of suitable size for X-ray diffraction multiple rounds of seeding are required. Suitable cryo-protection conditions have been developed allowing data collection at cryogenic temperatures. The crystals diffract beyond 3.0Å resolution when subject to synchrotron radiation and two native data sets have been collected with these crystals. A 98% complete data set to 3.5Å has been collected in-house together with a 92% complete set to 2.9Å at the E.S.R.F. In the search for isomorphous heavy atom derivatives, crystals soaked in K_2PtCl_6 looked promising judging from the statistics from SCALEPACK and SCALEIT. Unfortunately, the data were not good enough to locate heavy atom positions and hence solve the Patterson. Several metals have however been eliminated as potential derivatives. Early attempts to bind mercury and uranium isomorphously are encouraging and these avenues should be pursued further. A MAD experiment was also performed utilising platinum's ability to scatter X-rays anomalously. After the necessary data processing, heavy atom positions were located and refined to very low occupancy making phase determination impossible.

The protein has also been overexpressed in a minimal media (minus methionine) supplemented with seleno-methionine. Mass spectrometry confirms selenium incorporation at least 75% providing a route to structure solution through MAD. An 85% complete data set was collected using the selenomethionine variant crystals again at the E.S.R.F. Unfortunately heavy atom positions were indeterminable form this data.

The underlying problem throughout this project has been the inherent low diffracting power of the crystals obtained and the variability from batch to batch. The crystals are at best 0.06mm in the third dimension and diffract to low resolution in-house. The knock on effect of this has resulted in data collected in-house being more difficult to process and hence draw conclusions from. Perhaps the most disappointing aspect of this was the inability to characterise the large mosaic spread that was evident when experimenting with the crystals at the synchrotron. The percentage of glycerol used in cryoprotection and the protocol employed were modified without any observable decrease in mosaic spread. When mounting the crystal on the goniometer head loops just slightly bigger than the crystal itself where routinely used so to keep the scatter observed from the solvent to a minimum. However, loops so small may be inadvertently inducing stress on the crystals that in turn increases the observed mosaic spread. Conversely if loops of larger diameter were used this stress would be minimised and the crystal would be held in the centre of the loop by surface tension.

In nature it is known that plants and insects protect vital proteins from extreme conditions by wrapping them in a protective sugar cocoon (Potera, 1998). The sugars either forms a glass armour or vitrifies preventing the proteins from being damaged, then on coming into contact with water the sugars melt and the proteins regain their activity. This process may be directly applicable to the cryoprotection of protein crystals in X-ray crystallography. Cryoprotecting the crystals in a mixture of precipitant, glycerol and for example trehalose may prevent any internal damage to the crystal and subsequently reduce the observed mosaic spread.

Ultimately, if the orthorhombic crystal form can be reproduced then many problems may be solved simultaneously. This crystal diffracted to a higher resolution in-house (2.7Å), (grown without any optimisation of the crystallisation parameters), than any seeded crystals did at the synchrotron. Incorporating seleno-methionine into these crystals may provide a trouble free and quick route to structure solution. Alternatively, crystals from *K. pneumonia* have also been grown. Should optimisation of these conditions yield well diffracting crystals then structure solution via this route should be pursued. The *E. coli* structure could then be determined by molecular replacement using the native data set already collected at the E.S.R.F.

2.5 Future Work

The observed mosaic spread of the crystals will be assessed when the crystals are mounted in loops considerably larger than the crystals themselves. In addition, different sugars will be characterised as potential cryo-protection agents, again with a view to reducing the mosaic spread of the crystals. Should either of these alterations reduce the mosaic spread such that it becomes manageable then a structure may yet be obtained from these crystals. Failing these measures alternative crystallisation conditions for the *E. coli* enzyme may be required.

The *K. pneumonia* enzymes crystallisation conditions should be optimised further. These crystals may be of better quality than the *E. coli* variety and provide a quicker route to structure solution.

Chapter 3

The 1.85Å structure of succinylated concanavalin A bound to the synthetic bivalent ligand (1, 3-di(N-propyloxy-α-D-mannopyranosyl)-carbomyl 5-methylazido-benzene).

3.1 Summary

The low affinity binding that typifies protein-carbohydrate interactions is a primary concern for those striving to develop small molecule inhibitors of such interactions. Nature appears to overcome this problem through multivalency, where many simultaneous binding events overcome weak individual binding events. Accordingly, synthetic polyvalent ligands should access this effect. In collaboration with Eric Toone, Duke University, we have undertaken a project to determine the molecular basis of such interactions.

Crystals of x-ray diffraction quality have been obtained of succinyl con A bound to a synthetic bivalent ligand and the three-dimensional structure determined to 1.85Å by molecular replacement. Initially, diffraction data were recorded from a single crystal to 2.65Å at room temperature and the structure refined to this resolution. Subsequently, higher resolution data were collected on a flash cooled crystal at the Synchrotron Radiation Source on station PX9.6 Daresbury. The final model, refined to 1.85Å has good geometry and a final R factor of 20.8% and a R_{free} of 23.3%. The structure shows the protein crystallises as a tetramer cross linked to 4 other protein tetramers, creating infinite plains of tetrameric molecules. A dimer is observed in the asymmetric unit and the ligand spans two monomers. Both mannose residues are recognised at the so called monosaccharide binding site. While both amides remain planar in each structure, in the low resolution structure they adopt alternative conformations in *cis* and *trans,* breaking the two-fold rotation symmetry of the ligand. The interactions in the succinylated con A tetramer are considerably less than those observed in native con A, however they are extensive enough in that one would identify the protein to function as a tetramer. This is the first crystal structure of a protein bound to a synthetic multivalent carbohydrate ligand and together with data from agglutination assays,

titration micro-calorimetry and analytical ultracentrifugation has profound implications for the design of multivalent ligands.

The structure of a trivalent ligand bound to succinyl con A has also been determined. However, the third mannose residue is disordered and the resultant structure is identical to the bivalent ligand / succinyl con A complex.

3.2 Introduction

Proteins which recognise carbohydrates are ubiquitous and their functions diverse, (Sharon and Lis, 1989). They recognise carbohydrates with varying degrees of specificity and bind them with varying degrees of affinity. The processes, which govern such interactions, are not well characterised but the advent of more advanced biophysical techniques has made an understanding of such processes attainable.

Lectins comprise a varied family of sugar binding proteins and are found in all types of organism binding carbohydrates reversibly and with exquisite specificity (Goldstein and Portez, 1986). Plant lectins in particular are of intense interest as they are more tractable to characterisation by a broad range of biophysical techniques unlike their mammalian counterparts.

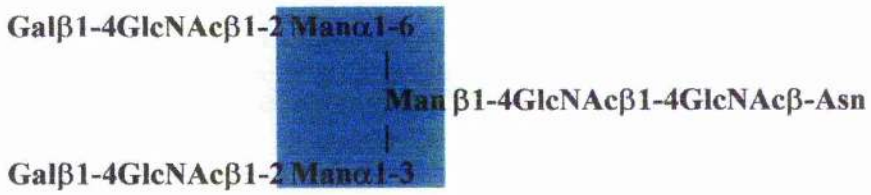
Concanavalin A (con A) is the most extensively studied member of the plant lectin family. The con A monomer consists of 237 amino acids and exists as a dimer in the pH range 5.0 – 5.6, while at higher pH tetramerisation is more common until at pH 7 tetramers are the predominant species (Goldstein and Poretz, 1986). Succinylation provides a form of the protein that is thought to be dimeric at all pH values (Gunther *et al,* 1973).

Con A was first isolated and crystallised from the Jack Bean in 1919 (Sumner, 1919). However, it was not until 1975 that the native X-ray structure was determined (Reeke, *et al,* 1975) and it was a further 14 years before the first con A carbohydrate complex (con A / α-D-mannopyranoside) was obtained (Derewenda *et al,* 1989). The delay in obtaining a complexed structure underlined the problems in obtaining suitable crystals of such complexes. A higher resolution study of this complex followed in 1994 (Naismith *et al,*

1994), allowing an extensive description of the monosaccharide binding site. The sugar is anchored to the protein by several direct hydrogen bonds, polar contacts and Van der Waals interactions burying 75% of the sugars accessible surface area. There are very limited main chain conformational changes upon carbohydrate binding to con A. Consequently, in relation to other carbohydrate binding proteins, more of the carbohydrate is exposed to the solvent and the full binding potential of the sugar is not realised (Quiocho, 1993).

In the years following this structure a plethora of new structures emerged revealing the specifics of the oligosaccharide binding site. Oligosaccharide binding is centered on a trimannoside core found in all *N*-linked glycans, Figure 3.1 (Rini, 1995). In 1996 Naismith and Field elucidated the crystal structure of con A bound to the trimannose highlighted in Figure 3.1 (Naismith & Field, 1996). The structure reveals the trisaccharide core sitting in an extended con A binding site with specific H-bonding between all three sugars and the protein. They also identified a structurally conserved water. The same water is present in the native structure and was proposed to be part of the binding site, Figure 3.2 (Naismith and Field, 1996).

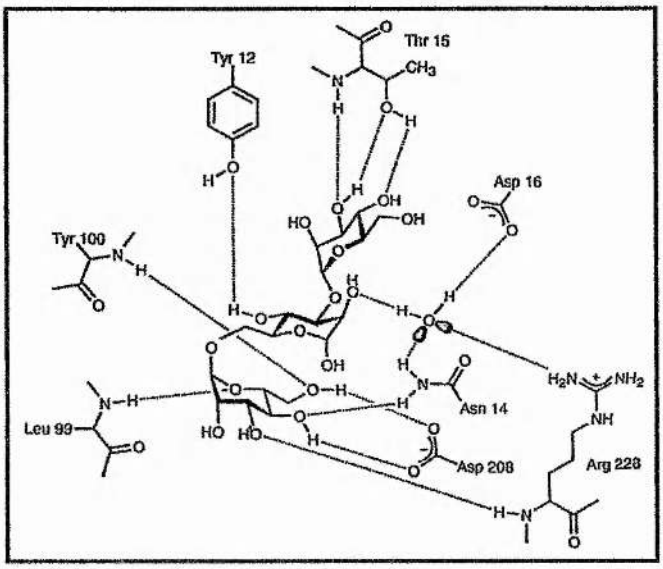
Figure 3.1: Biantennary N-linked glycan. The trimannoside core is shaded.



When an oligosaccharide is specifically recognised, it exhibits an increased affinity for binding over the cognate monosaccharide. This is most likely directly related to the

increased number of interactions possible between individual sugar rings and the extended binding site. The role water molecules play in this recognition is controversial. Originally it was proposed that water molecules are recruited to the binding site to mediate and stabilise interactions between the sugars and protein enhancing overall specificity. However, Naismith and Field's structure (1996) suggests only conserved waters already bound to the protein are significant in recognition.

Figure 3.2: Schematic representation of the hydrogen bonds between the sugar and the protein for the 2.0Å crystal structure of the trimannoside con A complex. The structurally conserved water is also displayed (Naismith and Field, 1996).



Ligand Multivalency

As therapeutic agents, carbohydrates are poor candidates. Oligosaccharides are too polar for satisfactory uptake and monosaccharides bind too weakly to be of value. However, it appears that in nature many carbohydrate binding proteins function as oligomeric species. This suggests that a possible solution to this conundrum may lie in the use of polyvalent ligands where multiple simultaneous binding events may overcome weak individual interactions. Several groups have pursued this, synthesising multivalent ligands and then assessing their affinity for carbohydrate binding proteins, reporting quite spectacular results. Enhanced affinities ranging from 10 to 10^9 over the appropriate monovalent ligand have apparently been observed (Page *et al*, 1996; Quesenberry *et al*, 1997; Ashton *et al*, 1997; Zanini and Roy, 1997; Sigal *et al*, 1996).

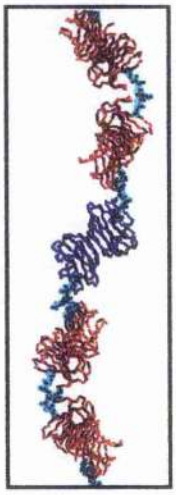
Several structures exist of lectins bound with naturally occurring multivalent saccharides, occurring as cross-linked monomers / dimers, forming linear arrays of protein – carbohydrate molecules (Bourne *et al*, 1994; Wright and Hester, 1996).

Bourne describes oligosaccharide bound structures of the bovine heart muscle lectin Galectin-1 bound with an octasaccharide and an asparaginyl-nonasaccharide (Bourne *et al*, 1994). Both oligosaccharides are biantennary and contain LacNAc at the non-reducing end. For each complex three crystal forms are obtained as a result of different low energy oligosaccharide conformations in solution. The structures reveal infinite chains of lectin dimers cross linked trough the LacNAc units of the oligosaccharides, Figure 3.3.

It appears that the formation of such lattices is dependant on the valence of one of the reactants being at least two, allowing cross-linking in two or three dimensions. Bhattacharyya and co-workers (1987), who characterised solutions containing a single lectin

in the presence of two oligosaccharides, have provided evidence for this. They discovered that only homogenous precipitates are formed and that heterogeneous complexes existed only in solution, that is they failed to precipitate. From this they concluded that each oligosaccharide formed a unique homogenous cross-linked lattice with the lectin and they were able to view these lattices by electron microscopy, revealing distinct patterns characteristic for the particular lectin carbohydrate complex (Bhattacharyya *et al*, 1987; Gupta *et al*, 1994).

Figure 3.3: Section of the infinite chains found in the crystal structure of bovine heart galectin-1 octasaccharide complex (Bourne et al, 1994).



Energetics of Protein – Carbohydrate Interactions

The principal method for assessing protein – carbohydrate interactions until recently has been the inhibition of hemagglutination or oligosaccharide precipitation. The protocol involves measuring the ability of a synthetic ligand to inhibit aggregation of a polyvalent ligand. Whilst this technique offers the advantage of simplicity, results can be dominated by kinetic factors making the thermodynamic contributions undeterminable. The aggregates

formed are ordered but not crystalline and structural data at the molecular level of the protein – ligand interaction is lacking. The advent of isothermal titration microcalorimetry allows the direct measurement of thermodynamic data. In this technique a soluble protein is titrated with aliquots of a soluble ligand. The heat evolved during ligand addition is deconvoluted to yield a binding constant K, which is related to the free energy of binding, $\Delta G^{\circ} = -RT \ln K$. In addition, binding enthalpies are measured directly and this allows the entropy of binding to be calculated via $\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$. Specific recognition of an oligosaccharide is indicated in a more negative free energy of binding compared to the cognate monosaccharide. Almost all carbohydrate interactions are entropically unfavourable in water (Chervenak *et al*, 1992; Mandal *et al*, 1994). Calorimetric analysis showed that a considerable proportion of the observed enthalpy increases arose from solvent reorganisation, that is a transfer of molecules between the bulk of the solvent and the solvation shell of the protein and the ligand (Chervenak and Toone 1994). This enthalpy – entropy trade off is a general feature of lectin carbohydrate binding (Grunwald and Steel, 1995). The binding of larger specific ligands has more favourable enthalpy terms but this is compensated for to some extent by an unfavourable entropic term. This is interpreted in terms of changes in the rotational degrees of freedom as torsion angles are frozen, and in terms of solvent reorganisation upon binding (Toone, 1994).

The 'Cluster Glycoside Effect'

As well as being important recognition molecules in cell wall synthesis, the onset of a variety of viral, parasitic, mycoplasmal and pathogenic infections may also be triggered by a protein – carbohydrate interaction (Dwek, 1996). The prevalence of these interactions has led to

intense research towards the development of carbohydrate based therapeutics (Lowe and Ward, 1997). Such species would provide a non-cytotoxic range of therapeutics applicable to a wide range of human diseases by interfering with the biological recognition process. However, the ubiquity of carbohydrates presents an immense challenge in the rational design of therapeutics, as the physical basis of protein – carbohydrate interactions is still poorly understood. In addition, the weak interaction energies associated with protein – carbohydrate interactions has delayed progress, typical binding constants ranging from millimolar to micromolar range (Mandal *et al*, 1994). An understanding of the recognition processes that govern such reactions is necessary. Recent years have seen significant progress made towards this through the combination of X-ray crystallography and isothermal micro-calorimetry.

Carbohydrate binding proteins (and carbohydrates) virtually always exist as oligomeric structures *in vivo,* suggesting nature overcomes the problem of weak carbohydrate / lectin binding through multivalency. That is, the use of polyvalent ligands can overcome the weak character of each individual interaction through simultaneous binding events. Hence, chemists have designed and synthesised polyvalent ligands, on the assumption that they would have high enough affinities to be useful therapeutic agents, bearing in mind monosaccharides bind too weakly, and oligosaccharides are too polar to be of general use. Several groups have pursued the goal of tight binding through the synthesis of a series of ligands of valencies ranging 2 through 20 for binding to lectins. Quite spectacular results have been observed with apparent binding affinities increasing from 10^2 to 10^9 over the cognate monosaccharide (Page *et al* 1996; Sigal *et al*, 1996; Ashton *et al*, 1997 and Zanini

and Roy, 1997). This striking enhancement in binding has led to the general acceptance of what is termed 'the cluster glycoside effect' (Meunier & Roy, 1996).

Despite the apparent success of the cluster glycoside effect, a molecular interpretation of the phenomenon has been difficult to ascertain. Typically they are explained by way of entropic gains, either an effective concentration argument or a chelate effect. More often than not the linker region between recognition epitopes is too short to bridge two binding sites on the one molecule. Hence, the apparent increase in affinity is a result of two separate molecules being bridged, Figure 3.4a, and not when a multivalent ligand binds at one molecule, Figure 3.4b, (a true chelate effect). Cluster glycoside effects have been observed for monovalent lectins (Miller *et al*, 1992).

Figures 3.4 a & b: Cartoon representation of a multivalent ligand (white) binding with a) Two lectins (orange) and b) One bivalent lectin (orange).

a)  **b)** 

Equation 1 relates the binding free energy for a bivalent ligand to a bivalent receptor, to the analogous monovalent binding free energies.

Equation 1: The binding free energy related to the analogous monovalent binding free energy where ΔG_m represents the binding free energy of the component of monovalent ligands and ΔG_i is an interaction energy (Jencks, 1981).

$$\Delta G^\circ = 2\Delta G^\circ_m + \Delta G^\circ_i$$

Assuming the linker domain is of sufficient length and flexibility that allows the recognition domains of the lectin to be in their most favourable orientation for binding, then enthalpy contributions to $\Delta G°_i$ will be small and entropic effects will dominate $\Delta G°_i$. The tethering of two binding sites will reduce the overall translational and rotational entropy of binding ($\Delta S^0 \downarrow$). During the complete binding of a bivalent lectin by monovalent ligands, three molecules are converted to one; during the corresponding binding of a bivalent ligand only two particles are converted to one. That is, the binding of a bivalent receptor proceeds with entropic savings equivalent to the translational and rotational entropy of one monovalent ligand. This is known in chemistry as the chelate effect and current estimates as to the translational and rotational energy barrier to biomolecular complex formation range from 2.5 to 15 kcal mol^{-1} at room temperature (Mandal and Brewer, 1992). Countering this effect, the introduction of a flexible linker domain introduces an entropic penalty relative to monovalent binding, as conformational degrees of freedom in the linker domain are 'frozen' out in binding. The magnitudes of these effects are difficult to ascertain, though current estimates are around 0.5kcal mol^{-1} per bond frozen during binding (Weis *et al*, 1996). As the distance between recognition domains on most polyvalent lectins is between 20Å and 70Å, it then seems unlikely that a chelate effect will be beneficial to reversible binding free energies, Table 3.1.

Table 3.1: Estimation of the entropy penalty (relative to monovalent) in bimolecular complex formation.

Distance Between Recognition Domains (Å)	Number of C-C Bonds	Entropy Penalty (kcal Mol^{-1})
20	**8**	**4**
50	**20**	**10**
70	**28**	**14**

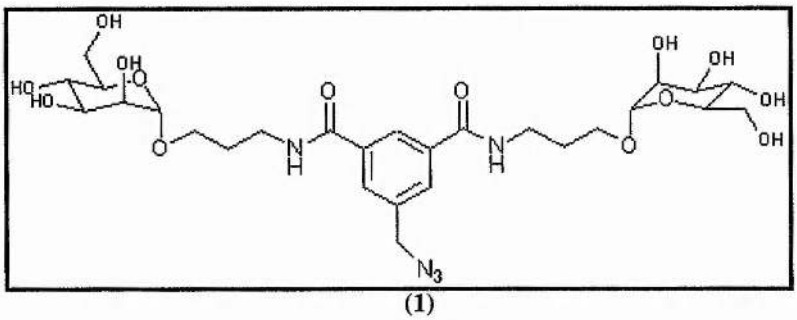
In collaboration with Eric Toone (Duke University, North Carolina, USA) a program designed to explicitly separate the thermodynamic contribution of polyvalent ligand binding to their apparent affinities measured by assays has been undertaken. In this chapter the crystal structure of succinylated con A complexed with the synthetic bivalent ligand (1) and trivalent ligand (2) Figures 3.5 and 3.8 are reported. Both ligands are designed as mimics of biological sugars. Titration microcalorimetric binding studies of ligand (1) with con A reveal a binding constant of 1.9×10^4 mol^{-1}, a binding enthalpy of -5.9 kcal mol^{-1} and a free energy of binding equivalent to –24 kJ mol^{-1}. These are comparable with those of α - D – mannopyranoside, binding constant 7.6×10^3 mol^{-1}, a binding enthalpy of –6.8 kcal mol^{-1} and free energy of binding equal to –22 kJ mol^{-1} (Chervenak & Toone, 1994). This work represents the first structural characterisation of a synthetic multivalent monosaccharide analogue complexed to a protein.

3.3 Experimental.

Crystallisation and Data Collection of the Bivalent Ligand 1 Complexed with Succinyl con A.

Succinylated con A was purchased from Sigma (Poole, United Kingdom). The bivalent ligand (1, 3-di (N-propyloxy-α-D-mannopyranosyl)-carbomyl 5-methlazido-benzene) was synthesised by our collaborator Eric Toone, Figure 3.5.

Figure 3.5: The bivalent ligand 1,3-di(N-propyloxy-α-mannopyranosyl)-carbomyl 5-methylazido-benzene.



(1)

Crystallisation trials were performed using the sitting-drop vapour diffusion method (Ducruix and Giege, 1992). A solution of succinylated con A (1.2mM) / ligand (18mM) was prepared in 20 mM TRIS pH 7.0, 100mM NaCl, 1mM $CaCl_2$ and 1mM $MnCl_2$. Co-crystals of the succinyl con A-bivalent ligand complex were obtained after an initial screen consisting of 10, 15, 20 and 25% PEG 6K at pH units 4.0, 5.0, 6.0, 7.0, 8.0 and 9.0. Small crystals (0.1 x 0.2 x 0.2mm) were initially obtained with a PEG concentration of 20%, pH 5.0. Optimisation of both the PEG and ligand concentration followed along with pH (Moothoo and Naismith, 1998a). Crystals of dimensions 1.0 x 0.8 x 0.6mm were obtained with the conditions outlined above but with a ligand concentration of 5mM. Crystal growth was complete within 14 days. In total 12 crystals were mounted in thin walled glass

capillaries and their diffraction limits tested by exposure to X-rays. The resolution limit varied markedly from 3.5Å to 2.65Å. Larger crystals did not improve the resolution.

All diffraction data were collected at 293.5 K from a single crystal (0.6 x 0.5 x 0.3mm) which diffracted to 2.65Å. Data was recorded on the Nonius/MacScience DIP2000 dual image plate and X-rays generated using an Enraf-Nonius FR951 rotating anode generator and focused using the MacScience mirror system, through a 0.5mm collimator. Data were collected as 92 non-overlapping images, 25 minute 1° oscillations with a crystal to detector distance of 140mm. The programs DENZO and SCALEPACK (Otwinowski, 1996) were used to index and merge the data. The crystal was indexed in a centered orthorhombic space group with unit cell dimensions a = 99.1Å, b = 127.4Å, c = 118.9Å. Analysis of diffraction data identified systematic absences consistent with space group $C222_1$. A dimer in the asymmetric unit gives rise to a V_m (Matthew's, 1968) of $3.8Å^3\ Da^{-1}$ indicating a solvent content over 65%. Data quality is summarised in Table 3.2. Davina Moothoo is acknowledged for practical assistance relating to the crystallisation and data collection reported here.

Table 3.2. Quality of the data for the 2.65Å model of succinyl con A bound to a synthetic bivalent ligand.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
26.00-5.69	2143	93.6	4.0	2.6
5.69-4.53	2206	99.6	4.5	3.0
4.53-3.96	2177	99.8	5.4	3.0
3.96-3.59	2171	99.8	6.5	3.0
3.59-3.34	2138	99.5	8.0	3.0
3.34-3.14	2151	99.2	10.3	3.0
3.14-2.98	2124	99.2	12.1	3.0
2.98-2.85	2112	99.1	15.6	3.0
2.85- 2.74	2074	96.7	19.2	3.0
2.74- 2.65	1886	87.6	20.7	2.9
26.00-2.65	**21182**	**97.4**	**7.2**	**2.9**

High Resolution Data Collection

Data were also collected from a larger cryoprotected (15% glycerol) crystal, 0.8 x 0.6 x 0.4mm, at a later date at the Daresbury Synchrotron Radiation Light Source on Station 9.6 using a ADSC quantum 4 CCD at a crystal to detector distance of 210mm. The wavelength was set to 0.87Å. The temperature for data collection was held at 100K. Data was recorded as 140 non-overlapping 2 minute 1° oscillations. The crystal had a mosaic spread of 0.45°. MOSFLM (Leslie, 1993) and SCALA (Evans, 1993) were used to index and merge the data. Data quality is summarised in Table 3.3.

Table 3.3. Quality of the data for the 1.85Å model of succinyl con A bound to a synthetic bivalent ligand.

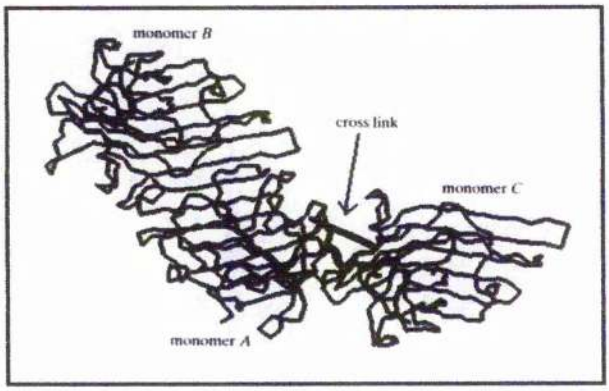
Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
25.0- 5.85	5254	87.1	7.5	3.0
5.85 – 4.14	3306	94.0	7.1	3.3
4.14 – 3.38	4335	96.0	7.5	3.4
3.38 – 2.92	5133	96.7	7.9	3.4
2.92 – 2.62	5805	96.9	8.6	3.4
2.62 – 2.39	6458	97.7	9.7	3.3
2.39 – 2.21	7035	98.2	11.4	3.3
2.07 – 2.07	7527	98.0	13.8	3.3
2.07 – 1.95	7992	97.9	18.4	3.1
1.95 – 1.85	5734	67.1	30.9	2.4
25.0 – 1.85	**55058**	**92.5**	**8.2**	**3.2**

Structure Solution

The structure was elucidated using the molecular replacement package as implemented in CCP4 (Collaborative Computational Project Number 4, 1994) program AMoRe (Navaza, 1994). The conventional con A dimer (monomer A and monomer B) from the trimannoside complex of conA (PDB code 1CVN) was used as a search model with all metals, waters and sugar atoms removed and remaining atoms set to full occupancy. A cross-rotation function gave only one solution $\alpha = 0.01$, $\beta = 66.59$, $\gamma = 313.91$ with a correlation coefficient of 0.28, 12 standard deviations above noise for the 10 - 4Å data. The translational search produced a final solution x = 0.1339, y = 0.2521, z = 0.2987, correlation coefficient of 0.84, 33 standard deviations above noise, R-factor 38%. A translational search in C222 gave no solution. At this point the space group was definitively assigned as $C222_1$. The asymmetric unit was

redefined such that one monomer of the con A dimer (monomer A) was linked via the ligand to another monomer (monomer C). Monomer C was generated by applying the symmetry transformation (½ - x, ½ - y, z + ½) to monomer B, Figure 3.6

Figure 3.6: The conventional dimer of con A is formed by monomer A and monomer B, the ligand crosslinks monomer A to monomer C. The cross linking ligand is shown as a straight line between the two sugar binding sites.



Refinement.

The solution was imported into CNS (Brunger *et al*, 1998) and refinement carried out using the free R-factor (R_{free}) as a guide. This was calculated on 10% of the data before any refinement had taken place and was subsequently excluded from all further refinement. Non – crystallographic restraints were applied to thermal and positional parameters throughout refinement. The initial free *R* factor was 38%. Rigid body refinement of monomers reduced this to 30%. Restrained refinement of positional and thermal parameters further reduced the R_{free} to 25%. At this stage $2F_o - F_c$ and $F_o - F_c$ maps were calculated and clear density was visible for metal ions ($2Mn^{2+}$ and $2Ca^{2+}$) and the ligand. The ligand geometry was defined using the Engh and Huber compendium of bond distances (Engh & Huber, 1991) and was included in the model for further refinement. Alternating cycles of automated refinement,

NCS restrained thermal and positional, with manual intervention using O proceeded smoothly. No cut-off's were applied to the data except the 10% of data removed to monitor refinement. Overall anisotropic thermal parameters and bulk solvent corrections were applied to the data using standard CNS protocols (Brunger, 1998) for refinement. Electron density maps were calculated in CNS and converted for viewing in O (Jones *et al*, 1991) with MAPMAN (Kleywegt, unpublished program).

Water molecules were added to the model in batches providing they satisfied the following criteria: (1) They corresponded to a peak > 3.5σ in the F_o - F_c map, (2) they made hydrogen bonds with reasonable stereo chemistry, (3) they reappeared in at least 1σ in subsequently calculated $2F_o - F_c$ maps and (4) a drop in R_{free} was observed. The progress of refinement is summarised in Table 3.4. The quality of the final 2.65Å model is shown in Table 3.5.

Table 3.4. Progress of refinement for the 2.65Å bivalent ligand – succinyl con A model.

Refinement	**R factor / R_{free}**	
Starting	34%	38%
CNS : Rigid body fitting of monomers (30 cycles)	26%	30%
O : Ligand, metal ions ($2Mn^{2+}$ and $2Ca^{2+}$) added and Waters added.		
CNS : Thermal (30 cycles) Positional (60 cycles)	22%	25%
O : More water molecules added		
CNS : Thermal (40 cycles) Positional (70 cycles)	18%	21%
O : More waters added.		
CNS : Thermal (40 cycles) Positional (40 cycles)	17%	20%

Table 3.5: Quality of the final model of the 2.65Å succinyl con A – bivalent ligand complex.

Data Collection		
Unique reflections		21182
Completeness of data(%)	(26.00 – 2.65Å / 2.75 – 2.65Å)	96.4 / 98.3
R_{merge} (I) (%)	(26.00 – 2.65Å / 2.75 – 2.65Å)	7.2 / 20.7
Average data redundancy	(26.00 – 2.65Å / 2.75 – 2.65Å)	2.5 / 2.2
% of data > 1σ	(26.00 – 2.65Å / 2.75 – 2.65Å)	94 / 83
Refinement		
Resolution range (Å)		26.00 - 2.65Å
R_{free} (%)		20.0%
R factor (%)		17.6%
Bond r.m.s. deviation (Å)[a]		0.008
Angle r.m.s. deviation (^{0})[a]		1.70
B-factor bounded atoms r.m.s deviation(Å^2)[c]		1.75
Ramachandran core/additional (%)[b]		86.7/12.8
Protein mean B-factor (Å^2)(all)		28.9

[a] r.m.s. deviation from Engh & Huber ideal values (Engh & Huber, 1991).
[b] Core and additionally allowed regions as defined by PROCHECK (Laskowski *et al*, 1993).
[c] B-factor deviation for bonded atoms

High Resolution Data

The refined 2.65Å ligand – succinyl con A complex co-ordinates were used as a model against which to refine the 1.85Å data. When the resolution is extended, it is important that if the R_{free} is to remain free, the indices that form the free-*R* set in the lower resolution data set (2.65Å) remain in the free-*R* set of the new high resolution set (1.85Å). This is accomplished by correct use of UNIQUEFY procedures in CCP4. Refinement proceeded using the same protocol used for the 2.65Å structure, rigid body followed by alternating

thermal and positional refinement incorporating manual intervention using O. The trend throughout refinement, lowering R_{free}, validates the refinement protocol.

3.4 Analysis of the 1.85Å Final Model

The final model contains a dimer of 546 residues, 2 Ca^{2+} ions and 2 Mn^{2+} ions, 330 water molecules and 160 sugar atoms. The quality of the data for the final model to 1.85Å is summarised in Table 3.6 and the Ramachandran plot (Ramakrishnan and Ramachandran, 1965), calculated using PROCHECK (Laskowski *et al*, 1993) for the final model is shown in Figure 3.7. Glycine residues (not restricted to any particular region of the plot) are represented by a ▲ and non-glycine residues by a ■. The darker the shaded region, the more favourable is the φ, ψ combination. No residues lie outside the additionally allowed regions. Regions are labelled as follows: A = core alpha, a = allowed alpha, ~a = generous alpha, B = core beta, b = allowed beta, ~b = generous beta, L = core left handed alpha, l = allowed left handed alpha, ~l = generous left handed alpha, p = allowed epsilon, ~p = generous epsilon.

Figure 3.7: Ramachandran plot for the 1.85Å succinyl con A – ligand model.

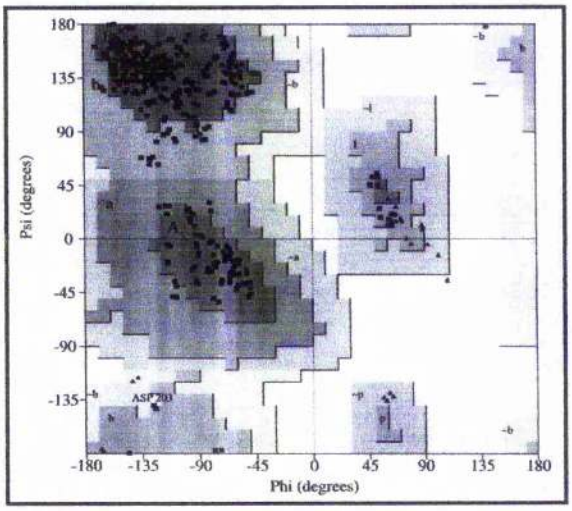


Table 3.6: Quality of the final model of the 1.85Å succinyl con A – bivalent ligand complex.

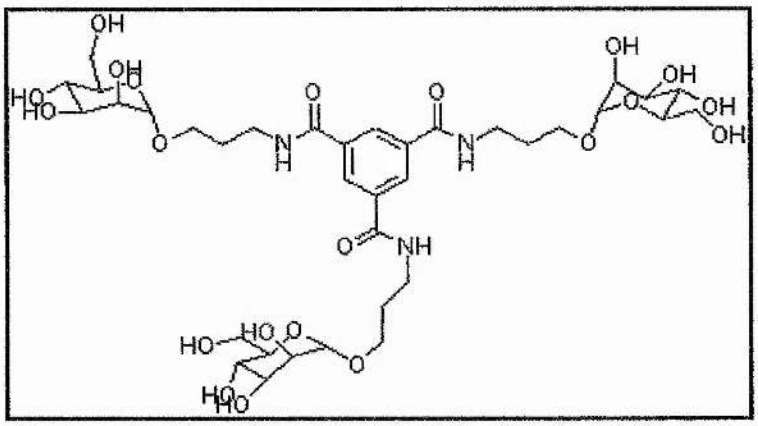
Data Collection		
Unique reflections		55473
Completeness of data(%)	(25.00 – 1.85Å / 1.95 – 1.85Å)	92.5 / 67.1
R_{merge} (I) (%)	(25.00 – 1.85Å / 1.95 – 1.85Å)	8.2 / 30.9
Average data redundancy	(25.00 – 1.85Å / 1.95 – 1.85Å)	3.2 / 2.4
Refinement		
Resolution range (Å)		26.00 – 1.85Å
R_{free} (%)		23.32%
R factor (%)		20.72%
Bond r.m.s. deviation (Å)[a]		0.01
Angle r.m.s. deviation (^{0})[a]		1.79
B-factor bounded atoms r.m.s deviation(Å^2)[c]		1.42
Ramachandran core/additional (%)[b]		86.5/13.3
Protein mean B-factor (Å^2)[c] (all)		34.6

[a] r.m.s. deviation from Engh & Huber ideal values (Engh & Huber, 1991).
[b] Core and additionally allowed regions as defined by PROCHECK (Laskowski *et al*, 1993).
[c] B-factor deviation for bonded atoms.

Structure Solution of Succinyl con A Bound to a Trivalent Ligand.

Crystals were also obtained of succinyl con A incubated with the trivalent ligand **2,** Figure 3.8.

Figure 3.8: The trivalent ligand crystallised with succinyl con A. R = H.



(2)

The same strategy for obtaining the initial crystals and then optimisation was followed as outlined for the bivalent complex. The best crystals grew after equilibration with a solution composed of succinylated con A (1.2mM) / ligand (18mM) prepared in 20 mM TRIS pH 7.0, 100mM NaCl, 1mM $CaCl_2$ and 1mM $MnCl_2$ and equilibrated against 0.1M Citric acid pH4.0, 18%PEG 6K. The crystals were pyramidal and the maximum size to which they grew was small relative to other con A carbohydrate crystals at 0.2 x 0.2 x 0.2mm.

The same protocol for data collection and structure solution was followed as outlined for the bivalent ligand **1**. The crystals diffracted in-house to 3.0Å. Indexing with DENZO (Otwinowski, 1996) indicated that the cell dimensions for the complex were almost identical to those for succinyl con A / bivalent ligand **1**, Table 3.7.

Table 3.7: A comparison of the unit cell and space group found for the succinyl con A bivalent ligand complex and succinyl con A crystallised with the trivalent ligand.

Succinyl con A / Bivalent Ligand	Succinyl con A / Trivalent Ligand
a = 99.1Å	a = 99.4Å
b = 127.4Å	b = 127.2Å
c = 118.9Å	c = 118.7Å
$\alpha = \beta = \gamma = 90^{\circ}$	$\alpha = \beta = \gamma = 90^{\circ}$
$C222_1$	$C222_1$

A data set to 3.0Å was collected on the trivalent succinyl con A complex. Molecular replacement and refinement proceeded in a manner identical to that for the bivalent complex. Clear electron density was evident at the monosaccharide binding sites for mannose residues, however no density was evident for the third sugar. It was concluded that this part of the ligand was disordered. To ensure no handling errors took place and that the bivalent ligand had not been crystallised and solved twice, the experiments (crystallisation and data collection) with the trivalent ligand were repeated. The initial results were reproduced. In addition, our collaborator confirmed by mass spectrometry that no degradation of the ligand had taken place.

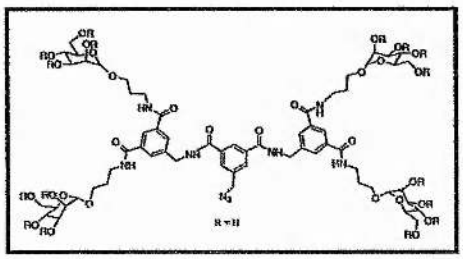
Crystallisation of Tetra and Hexa Valent Ligands with Succinyl Con A

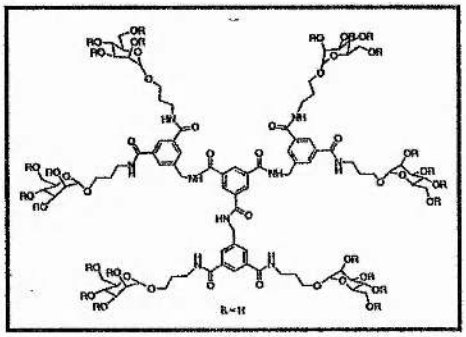
In addition to the bivalent (1) and trivalent ligand (2) two other con A multivalent ligand complexes were crystallised. Neither complex, as yet, has been characterised by x-rays.

The same crystallisation strategy as adopted for the other ligands was employed here. The tetravalent ligand, Figure 3.9, crystallised when prepared in a solution of 1.2mM succinyl con A / 9mM sugar after equilibration against 12 % PEG6K, 0.1M citric acid pH 5. Similarly the hexavalent ligand, Figure 3.9, crystallised when prepared in a solution of

1.2mM succinyl con A / 6mM sugar after equilibration against 8 % PEG6K, 0.1M citric acid pH 5.

Figure 3.9: The tetravalent and hexavalent ligands crystallised with succinyl con A.





Dynamic Light Scattering

Con A is reported to exist predominately as a dimeric species in solution at pH 5, whereas at neutral pH, the dimers dimerise, yielding a tetrameric species (Goldstein & Poretz, 1986). It has been reported that succinylation of the protein provides a dimer that remains dimeric at all pH levels (Gunther *et al*, 1973). At neutral pH tetrameric con A visibly precipitates from solution upon addition of ligand **1**. However, if succinyl con A is used in place of native con A no precipitation is observed.

To further explore the range of species present formed between the bivalent ligand and succinyl con A, mixed solutions were examined by dynamic light scattering. Solutions were prepared at 5mg ml^{-1} in 20mM TRIS pH 7, 10mM NaCl, 1mM $MnCl_2$, 1mM $CaCl_2$. Tetrameric con A was also analysed for comparison purposes. All solutions were 0.1μm filtered immediately prior to analysis and the results are shown in Tables 3.8, 3.9 & 3.10.

This experiment was repeated with the other ligands discussed in this chapter (tri, tetra and hexavalent) yielding essentially identical results to those obtained with the succinyl con A bivalent ligand complex.

Table 3.8: Analysis of native con A by dynamic light scattering. The average polydispersity value as a function of the hydrodynamic radius is 19%.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDa	Temp. ^{0}C	Count Rate	Base Line	%Radius / PolyD
1	0.833	578	4.46	1.1	108	27.9	432	1.000	24
2	0.846	581	4.45	0.4	108	27.8	432	1.000	9
3	0.838	562	4.60	1.4	117	27.9	452	1.007	30
4	0.842	541	4.79	1.5	129	27.8	457	1.009	31
5	0.846	569	4.53	0.3	112	27.9	442	0.009	6
6	0.846	575	4.50	1.3	111	27.8	439	1.000	28
7	0.850	592	4.36	1.0	103	27.8	438	1.001	23
8	0.844	563	4.60	0.6	116	27.9	441	1.000	13
9	0.844	575	4.50	0.7	111	27.9	438	1.000	15
10	0.840	572	4.52	0.8	112	27.8	442	1.000	17

Table 3.9. Analysis of succinyl con A by dynamic light scattering. The average polydispersity value as a function of the hydrodynamic radius is 42%.

	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDa	Temp. ^{0}C	Count Rate	Base Line	%Radius / PolyD
1	0.810	489	4.3	1.6	100	21.2	280	1.005	39
2	0.800	432	4.8	2.9	135	21.5	243	1.027	55
3	0.802	479	4.4	1.7	106	21.4	330	1.004	39
4	0.809	485	4.3	1.7	104	21.0	347	1.005	39
5	0.813	482	4.3	1.7	104	21.3	335	1.006	39
6	0.809	486	4.3	1.7	104	21.5	331	1.005	41
7	0.729	477	4.3	1.7	104	21.5	315	1.006	40
8	0.801	487	4.4	1.7	110	21.4	312	1.005	39
9	0.820	458	4.3	2.2	121	21.4	315	1.007	51
10	0.824	453	4.4	2.0	117	21.4	316	1.007	45

Table 3.10: Analysis of succinyl con A/bivalent ligand by dynamic light scattering. The average polydispersity Value as a function of the hydrodynamic radius is 96%.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDa	Temp. ^{0}C	Count Rate	Base Line	%Radius / PolyD
1	0.713	75	29.0	27.8	10200	21.2	280	1.090	95
2	0.725	107	20.6	19.8	4446	21.5	243	1.071	96
3	0.714	78	27.9	26.8	9306	21.4	330	1.080	96
4	0.725	76	28.5	27.4	9795	21.0	347	1.094	96
5	0.729	84	26.1	25.1	7905	21.3	335	1.087	96
6	0.725	85	25.7	24.7	7618	21.5	33	1.074	96
7	0.729	93	23.7	22.7	6237	21.5	315	1.074	95
8	0.735	91	24.2	23.2	6570	21.4	312	1.066	95
9	0.736	88	24.7	23.7	6942	21.4	315	1.064	96
10	0.735	89	24.8	23.6	6950	21.4	327	1.063	95

3.5 Discussion

The crystal structure of succinylated con A bound to the synthetic polyvalent ligand **1** has been determined to 1.85Å. The overall structure of the protein is essentially unchanged from previous con A complexes (Naismith *et al*, 1994; Naismith and Field, 1996; Moothoo and Naismith, 1998b)

Sugar Binding

Clear electron density is visible for the entire ligand in both the low and high resolution structures. However, at 2.65Å the carbonyl groups adopt different positions with respect to the aromatic ring, Figure 3.10. While both amides remain planar, one is orientated in a *cis*-conformation and the other in a *trans*-conformation. The ligand breaks its 2 fold rotation axis of symmetry, Figure 3.5, and each half of the ligand is chemically distinct in the structure. At high resolution both amides are in the *trans* orientation. The flexible portion of the ligand adopts a mainly extended conformation and separates the sugars by over 13Å, Figure 3.11. Both mannose residues are bound at the monosaccharide binding site of con A. The interactions between the ligand and the protein are shown in Table 3.11. The ligand is anchored to the protein by a combination of hydrogen bonds and Van der Waals interactions. The mannose residue dominates the contacts and the ligand and makes a few additional interactions with the protein. The orientation of the mannose residue within the binding site is little changed from that described previously for other con A carbohydrate complexes (Naismith and Field, 1996; Moothoo and Naismith, 1998b). The portion of the ligand outside the binding site occludes significantly more surface area than the simple monosaccharide, Table 3.11. Each mannose appears to be acting independently of the other in binding to the monosaccharide site. Modelling shows that maintaining the ligands strict

two fold rotation axis allows the protein to cross-link but such an array has no protein protein contacts and would appear unlikely to form crystals. Whether such a symmetrical arrangement exists in solution is unknown. The protein protein interaction between the linked monomers is small, burying only 318Å^2 of polar and 313Å^2 of apolar surface. Both values are commonly associated with crystal contacts. Hence, it appears that protein protein contacts are relatively unimportant in controlling the aggregation of con A by this ligand but may influence the conformation of the ligand.

Table 3.11: The interaction between one protein monomer and the ligand. The values for the mannose are shown in parenthesis. The mannose values are comparable to that observed in the ManOMe con A complex (Naismith et al,1994).

Van der Waals Contacts (<4.0Å)	48 (44)
Hydrogen Bonds and Polar Contacts	11 (10)
Polar Buried Surface Area	280 (216) Å^2
Apolar Buried Surface Area	237 (125) Å^2

Figure 3.10: F_o - F_c map contoured at 3.2σ with the refined ligand at 2.65Å. The phases for the map were calculated from the model at R_{free} 25% which did not include the ligand, metals or water molecules.

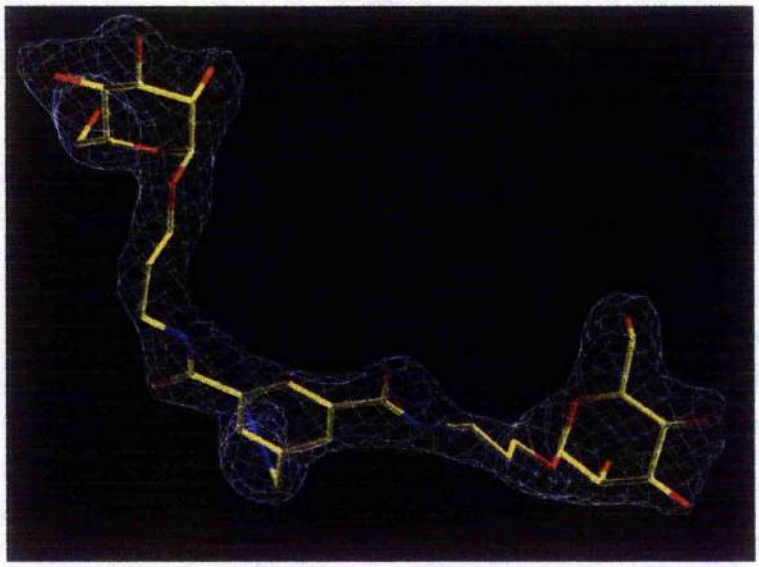
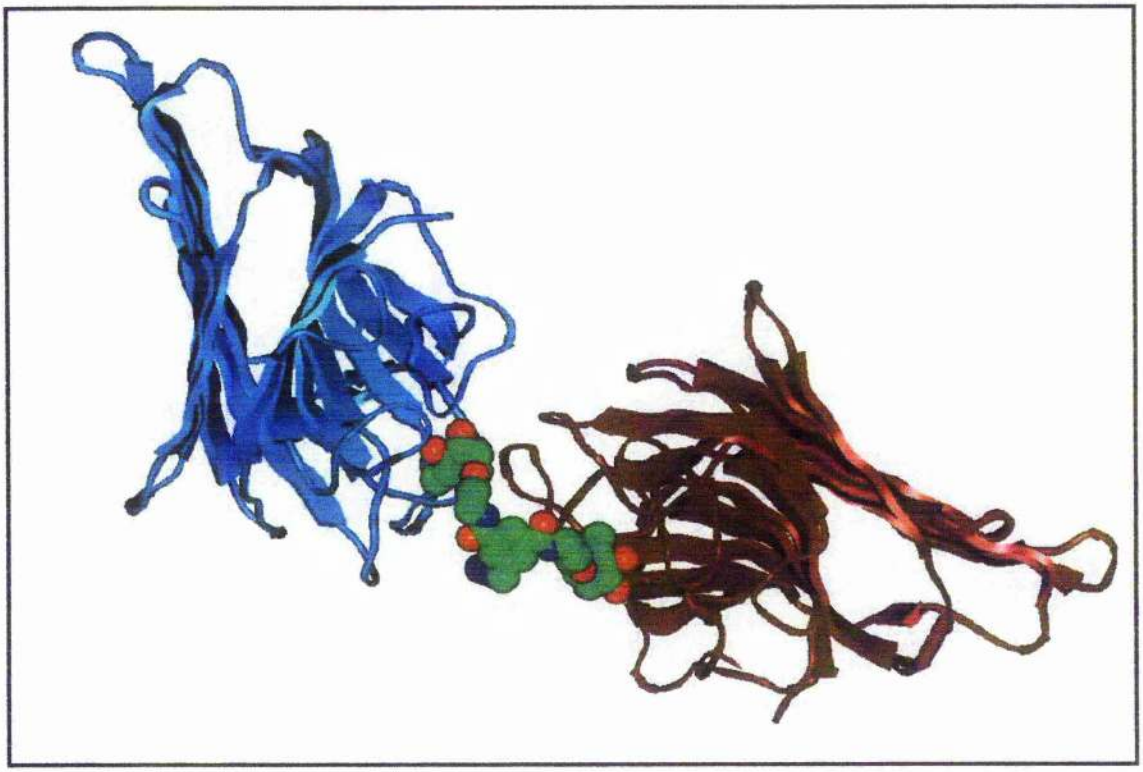


Figure 3.11: The ligand binds two monomers of succinyl con A effectively cross linking them.



Tetramer Organisation

Dimerisation of the dimers occludes 4300Å^2 of surface area. The buried surface area is considerably less than that found in native con A (5200Å^2) (Deacon *et al*, 1997; Reeke *et al*, 1975) and in the mannose con A complex (4800Å^2) (Naismith *et al*, 1994). There are 120 protein protein contacts less than 4.0Å between two dimers, compared to 260 for the native and 160 for the mannose complex. Although this tetramer is less tightly packed than the other two, one would still identify the protein to be a functional tetramer. The results observed from light scattering lend credence to this. Analysis of succinyl con A suggests a molecular weight of around 110 kDa, which is almost the exact weight of the predicted tetramer. The observation that succinylated con A forms into apparently normal tetramers

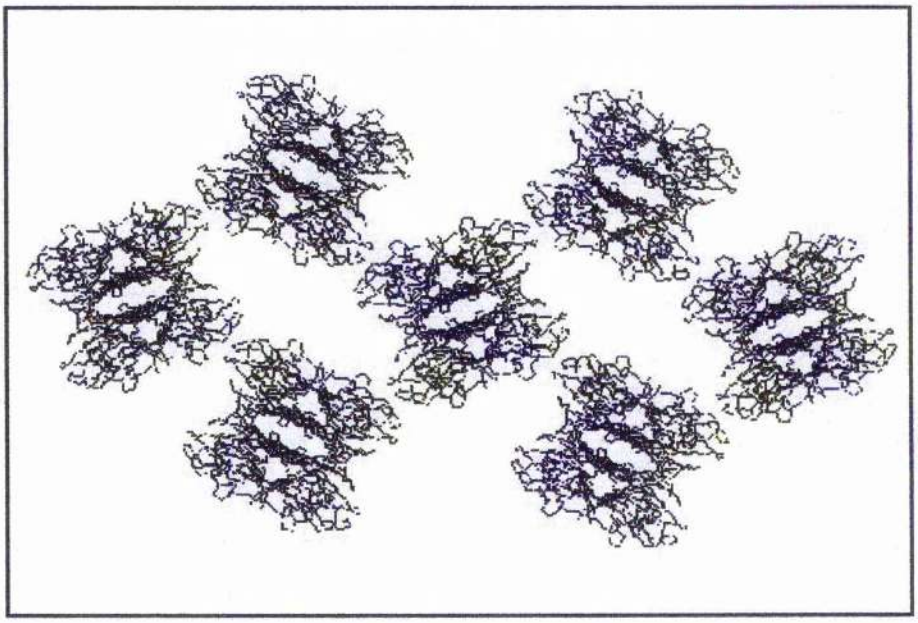
also requires caution when assuming that succinylated protein exists 'exclusively' in dimeric form.

Supramolecular Structure

The estimated molecular weight of the bivalent complex from light scattering ranges from 4456 kDa to 10200 kDa. That the profile is so broad suggests a range of multimeric species. Similar results were obtained when either the tri, tetra or hexavalent ligand was used in place of the bivalent ligand.

This phenomenon is supported in the crystal structure, as each tetramer is cross-linked to four other tetramers by the bivalent ligand. The interplay of crystal symmetry is complex and results in the formation of an infinite sheet of cross-linked tetramers, Figure 3.12. Two sheets pass through the unit cell and are related by a two-fold rotation axis.

Figure 3.12: The infinite sheet of succinyl con A tetramers found in the crystal. Each tetramer is bound to four other tetramers.



As the addition of ligand 1 to con A at neutral pH results in a precipitated solution then the structure reported here may reflect a more ordered array of this precipitate. (At pH 5 succinyl con A does not precipitate ligand 1). Correlation of this suggests the combination of protein and ligand valency is crucial in determining whether aggregate formation is energetically favourable. Where dimeric con A requires ligands with a valency of 3 or more to induce aggregation, tetrameric con A forms aggregates with the bivalent ligand. Apparently, aggregates only form under conditions where one dimer is linked to more than two other dimers. This crystal structure shows one dimer linked to two other dimers, the light scattering experiments similarly suggest a tetramer.

The results caution against an interpretation of solution behaviour based on observations from single techniques (aggregation states of con A are based on ultra-centrifugation experiments from the 1970's). Our results on the nature of succinylated con A suggest a much more dynamic environment than those based on the assumption that the solutions are composed of a single species. Combining results from light scattering and crystallography suggest a very complex equilibrium exists in solution between tetrameric and dimeric con A even after succinylation. Precipitation is a kinetically controlled process and caution as to how it is interpreted must be exercised. On the basis of these results it seems more likely that the ligands promote infinite arrays, however the rate of aggregate formation is highly dependent on the rates of exchange between dimers and tetramers. This rate of exchange is probably altered by pH and succinylation (hence the observed polydispersity of succinyl con A). Other lectin / carbohydrate complexes are known to exist as similar infinite arrays but as yet such aggregates have not been observed in solution.

Trivalent Complex

The trivalent ligand con A complex crystallised in the same space group as the bivalent complex with almost identical cell dimensions. The structure was solved to 3.0Å, however on visualisation no density was observed for the third mannose residue as all the monosaccharide binding sites were occupied. It appears likely that as a result of this the third sugar is disordered.

Data from calorimetric titration experiments show all ligands exhibit a 1:1 carbohydrate: binding site stoichiometry with the exception of the trivalent ligand **2** (binding stoichiometry 0.8) used here in crystallisation studies (Eric Toone, personal communication). However, the findings from this crystal structure raise the possibility that the soluble aggregates formed in solution pack in a similar fashion to the insoluble aggregates observed by X-ray diffraction. Theoretically, the binding of two carbohydrates by a trivalent ligand gives a binding stoichiometry of 0.67, that in practise it is 0.8 suggests that the aggregates are likely of lower order than the crystal form. However, the low stoichiometry is more likely as a result of some form of local order in aggregate formation that inhibits binding of the third saccharide of a trivalent ligand, even in solution.

3.6 Conclusions

Despite evidence to the contrary succinyl con A crystallises as native con A; in a tetrameric form. The bivalent ligand **1** infinitely cross-links monomers resulting in an infinite array of cross-linked tetramers. The sugars are bound at the monosaccharide binding site through a combination of hydrogen bonds and van der Waals interactions. The linking region of the ligand makes few additional contacts with the protein. Although the interactions holding the tetramer together are not as extensive as found in other con A complexes, by a standard set of crystallographic criteria one would still identify the protein to function as a tetramer. It appears that the energy barrier between dimers and tetramers is small enough that crystal packing forces are sufficient to drive formation of the tetramer (if dimers do exist in solution).

The results here also call into question the validity of the binding constants of multivalent ligands obtained by agglutination assays. This structure reveals the protein carbohydrate interactions observed with polyvalent ligands are essentially identical to those involved in the binding of α-methyl mannose. Assuming other multivalent ligands bind similarly then the enhanced binding efficiencies observed by assay are more likely kinetic in origin and result from aggregate formation and or protein precipitation. Such effects may or may not be relevant to the design of therapeutically useful saccharides. The conclusions of this work raise serious questions regarding the validity of the techniques and conclusions hence drawn from previous studies probing protein carbohydrate interactions.

3.7 Future Work

As this is the first structure lectin / multivalent ligand structure to be characterised in this manner, conformation of the results by characterisation of similar solutions would be valuable.

In addition throughout the work described in this chapter it has been assumed that the linker region between binding epitopes has been too short to span separate recognition domains on one molecule. Knowledge as to whether a true effect (increase in binding avidity) would be noticed if the linker region was of the sufficient length to permit multiple binding to one lectin is lacking. A crystal structure with such a ligand would identify such an entity, which would be useful in assessing true binding affinities where macroscopic aggregation prevails.

Chapter 4

The 2.7Å structure of concanavalin A complexed with the branched pentasaccharide

α-Man-(1→6)-[α-Man-(1→3)-α-Man(1→6)]-[α-Man-(1→3)]-Man.

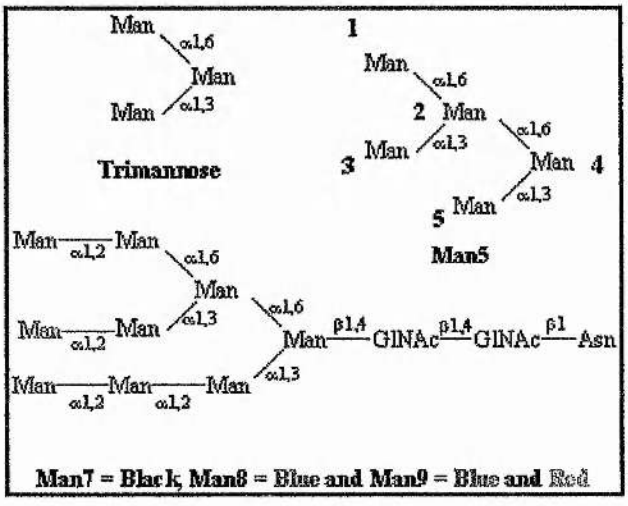
4.1 Summary

The structure of concanavalin A bound to the high order oligosaccharide α-Man-(1→6)-[α-Man-(1→3)-α-Man(1→6)]-[α-Man-(1→3)]-Man has been determined and refined at 2.7Å. Crystals of the complex are of dimensions a=81.73Å, b=66.12Å, c=108.66Å, β=97.69° and grow in space group $P2_1$. One tetramer is present in the asymmetric unit and each monomer binds one pentasaccharide molecule along with one manganese and calcium ion. In each subunit electron density exists for 4 of the 5 sugar residues in a well-ordered binding site and for the 5th sugar in 2 of the subunits. The pentasaccharide binds at the extended binding site of concanavalin A via the trimannose moiety of the molecule, stabilising the oligosaccharide through a combination of hydrogen bonding and Van der Waals interactions. Neither of the 4th or 5th sugar residues are specifically recognised. The pentamannose ligand does not crosslink to another concanavalin A, subunit as the 4th and 5th sugars adopt a conformation in which the sugar residue binds to the protein through a bridging water.

4.2 Introduction

Oligosaccharide chains of glycoproteins have been implicated as receptors in a variety of cellular recognition processes (Rini, 1995). They are found as sophisticated arrays on cell surfaces and are recognised with varying degrees of specificity by, among others, enzymes involved in cell wall synthesis. The asparagine linked or N-linked oligosaccharides form one class of these oligosaccharides which themselves can be further divided into three classes (Kornfeld and Kornfeld, 1985) and include the 'high mannoses'. N-linked high mannoses are branched chains of multivalent oligosaccharides of the type in Figure 4.1. The binding of multivalent lectins to N-linked high mannoses on the surfaces of cells can lead to the cross-linking and aggregation of certain glycolipids, which in turn is associated with a variety of biological responses including the molecular sorting of glycoproteins in the secretory pathways of cell (Chung et al, 1989) and the signal transduction mechanism of certain glycoprotein hormones (Sairam, et al 1989).

Figure 4.1: The high order mannoses discussed in this chapter and the Man5 used in this study. The mannoses of Man5 are labelled 1 – 5 for reference.



Higher order oligosaccharides binding to protein falls into one of two categories. They are either highly water mediated where the rearrangement of the water structure in and around the binding site mediates binding (Bourne *et al*, 1994). Or, interaction is through an extended carbohydrate binding site on the surface of the protein (Naismith and Field, 1996; Moothoo and Naismith, 1998b; Moothoo *et al*, 1999). The concanavalin A (con A) studies have employed high affinity oligosaccharides and provided a rationale for the high affinity displayed by con A for the trimannoside.

As the majority of the high mannose oligosaccharides precipitate tetrameric con A (suggesting cross-linking is apparent) the thermodynamic experiments to elucidate the binding affinities have primarily been performed with dimeric con A (succinyl conA) in titration microcalorimetry experiments (Mandal *et al*, 1994). The thermodynamic parameters for the binding of the pentamannose (Man5) used in this study, Figure 4.1, with succinyl con A are given in Table 4.1, along with those for the corresponding binding of the trimannoside, Figure 4.1, and higher order oligosaccharides (Man 7, 8 & 9), Figure 4.1, for comparison. The values reported for Man5 are consistent with a 1:1 binding stoichiometry where one mole of saccharide is bound per mole of lectin, whilst the higher order saccharides have binding ratios of around 0.92 suggesting some cross-linking. It should be noted that although the affinity displayed by succinyl con A for the 'higher' order oligosaccharides Man 7, 8 & 9 is essentially the same as that of the trimannose (and Man5), these glycopeptides have been reported to have nearly 7-fold higher affinities for tetrameric con A relative to the trimannose (Mandal & Brewer, 1993). However, these results are based on the same hemaglutination assays discussed in Chapter 3. The results are dominated by kinetic effects making the thermodynamic parameters indeterminable. That these oligosaccharides

apparently display greater affinities for tetrameric con A may be because the molecules are themselves of sufficient size that they have the ability to span two subunits and cross-link adjacent domains.

Table 4.1: Thermodynamic data for succinyl con A binding to a variety of high order oligosaccharides (Mandal et al, 1994).

	$-\Delta H°$ kcal mol^{-1}	$-T\Delta S°$ kcal mol^{-1}	$-\Delta G°$	$K_a(M^{-1})$
Trimannoside	14.3 (± 0.1)	6.5	7.8 (± 0.1)	5.3 (±0.22)x10^5
Man5	14.5 (±0.1)	6.6	7.9 (± 0.1)	6.6 (±0.40)x10^5
Man7&8	13.6 (±0.2)	5.6	8.0 (± 0.1)	7.7 (±0.12)x10^5
Man9	17.1 (± 0.4)	9.0	8.1 (± 0.1)	8.9 (±0.14)x10^5

As the results from the microcalorimetry suggest Man5 is recognised by con A with the same affinity as the trimannose, it is proposed the pentasaccharide will bind to con A via the trimannosyl moiety located on the α (1,6) arm of the molecule (sugars 1, 2, and 3). The thermodynamic parameters for the binding of Man5 to acetyl con A, Table 4.2, are in somewhat contrast to these results and demonstrate a binding stoichiometry approaching 0.5. This suggests a binding ratio of 1:2 for ligand to lectin, which would be consistent with crosslinking of con A monomers. The binding of both the α (1,6) arm (trimannosyl moiety) and the α (1,3) arm leads to a greater enthalpy change per mole of bound saccharide than that observed for the trimannose alone, and explains the greater $\Delta H°$ observed here relative to succinyl con A. The Man9 oligosaccharides display similar binding properties with acetyl-con A as the Man5, only demonstrating an even greater tendency for crosslinking relative to the analogous Man9 - succinyl con A experiment (Mandal *et al*, 1994).

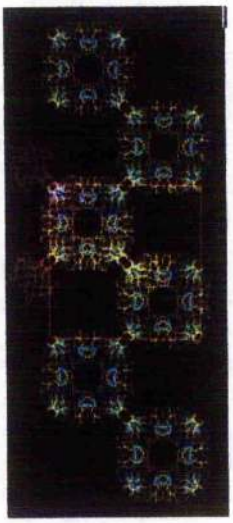
Table 4.2: Thermodynamic data for the pentamannose binding to acetyl-con A (Mandal et al, 1994).

	$-\Delta H°$ kcal mol^{-1}	$-T\Delta S°$ kcal mol^{-1}	$-\Delta G°$	$K_a(M^{-1})$
Man5	**18.4**	**10.8**	**7.6**	**$3.51(\pm 0.27)\times 10^5$**

A similar enthalpy difference is reported by Ambrosino and co-workers (1987) who, using titration microcalorimetry, suggest the enthalpy of binding of Man5 to dimeric con A is –18.8 kcal mol^{-1}, along with a binding stoichiometry of 0.5, implying the saccharide has the ability to bind two con A monomers.

The structure of the bulb lectin *Galanthus nivalis* agglutinin (GNA), complexed with Man5 has previously been solved to 2.0Å (Wright and Hester, 1996). Monomers of the GNA each contain three binding sites as a result of the 3-fold internal repeats that constitute the molecule. In the structure with Man5 two unique modes of binding are observed. At CRD1 (conserved recognition domain 1), the highest affinity monosaccharide binding site, Man5 crosslinks GNA dimers creating a pore like super structure, Figure 4.2. However, only 3 of the mannose residues (essentially the trimannose) are resolved in the structure and neither sugars 4 or 5 are visualised.

Figure 4.2: Pore-like assembly created by cross-linking of GNA dimers (Wright and Hester, 1996).



From the structural studies on the bivalent ligand discussed in Chapter 3, it is possible the Man5 ligand may crosslink con A subunits, though some of the microcalorimetric results suggest otherwise. A three dimensional structure will clear up this ambiguity and answer other questions. If the ligand does not crosslink subunits, then why not? Does specific recognition of the 4^{th} and 5^{th} sugars preclude crosslinking? And if so, then why is the sugar recognised with the same affinity as the trimannose?

In order to delineate the interactions involved in the binding of Man5 with con A, the three dimensional structure of con A complexed with this ligand has been elucidated and are discussed in this chapter.

4.3 Experimental

Crystallisation and Data Collection

Co - crystals of the con A – Man5 complex were initially obtained from a screen consisting off 10, 15, 20 and 25 % PEG 6K, pH 4, 5, 6, 7, 8 and 9. Optimisation of pH and PEG concentration followed (Moothoo and Naismith, 1998a). The largest crystals grew from a 5µl solution containing 0.6mM con A (Sigma, Poole, UK), 18mM α-Man-(1→6)-[α-Man-(1→3)-α-Man(1→6)]-[α-Man-(1→3)]-Man (Dextra Laboratories, Reading, UK), 1mM $MnCl_2$, 1mM $CaCl_2$, 20mM TRIS pH 7.0 and 0.1M NaCl mixed with 5µl of a reservoir containing 12% PEG 6K pH 5.0 in a sitting drop tray (Douglas Instruments) after 24 hours equilibration. Crystals grew as thin diamonds in dimensions ranging from 0.2 x 0.15 x 0.05mm^3 to 1.0 x 0.8 x 0.1mm^3.

To allow data collection at cryogenic temperatures, crystals were allowed to equilibrate in precipitant solution poisoned with 15% glycerol for 30 seconds immediately prior to data collection. All data were collected from a single cryo-protected crystal at 110K using the Enraf-Nonius/MacScience DIP2000 dual image plate. Data collection proceeded as 129 non-overlapping 20 minute 1° oscillations with a crystal to detector distance of 160mm. The data were indexed and processed with DENZO and SCALEPACK (Otwinowski, 1996) in a primitive monoclinic spacegroup. Systematic absences were consistent with a 2_1 screw axis and the space group was assigned $P2_1$ with unit cell dimensions a=81.73Å, b=66.12Å, c=108.66Å, β=97.69°. The asymmetric unit contains one tetramer; Matthews number 2.91Da $Å^{-3}$, approximately 57.39% solvent. The quality of the data is shown in Table 4.3.

Table 4.3. Quality of the data for the 2.7Å model of the con A pentasaccharide complex.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
25.00 – 5.80	2666	81.0	8.8	1.5
5.80 – 4.61	2725	84.8	10.5	1.9
4.61 – 4.03	2744	85.6	11.7	1.9
4.03 – 3.66	2722	85.7	13.7	1.9
3.66 – 3.40	2716	85.7	16.7	1.9
3.40 – 3.20	2727	85.9	21.3	1.9
3.20 – 3.04	2678	84.9	27.6	1.8
3.04 – 2.91	2660	84.0	34.0	1.8
2.91 – 2.80	2674	84.5	47.1	1.7
2.80 – 2.70	2587	82.4	52.1	1.7
25.00 – 2.70	**26899**	**84.4**	**15.3**	**1.8**

Structure Solution

The structure was determined using the molecular replacement package AMoRe (Navaza, 1994) as implemented in CCP4 (Collaborative Computational Project Number 4). The trimannoside con A complex tetramer (PDB Accession Number 1CVN) (Naismith and Field, 1996) was used as a search model. All metals, waters and sugar atoms were removed and the remaining atoms set to full occupancy. A cross rotational function gave one solution α = 89.02, β = 84.82, γ = 293.09, 19 standard deviations above noise. The translational search found one tetramer, solution x = 0.2727, z = 0.2411, yielding a correlation coefficient of 0.73 and an R-factor of 33.7%. Y is arbitrary in this space group.

Refinement

The solution was imported into CNS (Brunger *et al*, 1998) and refinement carried out using the free R-factor (R_{free}) as a guide. R_{free} was calculated on 5% of the data, which was excluded from all refinement calculations, providing an unbiased assessment of the progress being made. Non – crystallographic restraints were applied to thermal and positional parameters throughout refinement. The initial model for refinement included structurally conserved metals and waters consistent with other con A crystal structures; total 8 metals and 16 waters. A *cis* peptide was restrained between residues 207 and 208. The *R* and free *R* factors from the raw molecular replacement solution were 33.7 and 34.4% respectively. Rigid body refinement of the whole tetramer reduced these both to 30.5%. Further rigid body refinement of the individual monomers saw the *R* and R_{free} fall a further 0.3 and 0.1 % respectively. At this stage $2F_o$ - F_c and F_o - F_c maps were calculated. Clear density was visible for the three sugar residues at the trimannose binding site and a fourth sugar α(1-6) linked to the reducing mannose. These residues were included in the model for further refinement. The ligand geometry was defined using the Engh and Huber compendium of bond distances (Engh & Huber, 1991). Alternating cycles of refinement, NCS restrained positional and thermal parameters, followed with manual intervention using O, reducing the R_{free} to 27.9%. Water molecules were added to the model providing they satisfied four criteria: (1) They corresponded to a peak > 4.0σ in the F_o - F_c map, (2) they reappeared in at least 1σ in subsequently calculated $2F_o$ – F_c maps (3) they made hydrogen bonds with reasonable stereo-chemistry and (4) a drop in R_{free} was observed. No cut-off's were applied to the data except the 5% of data removed to monitor refinement. Overall anisotropic thermal parameters and bulk solvent corrections were applied to the data using standard CNS

protocols (Brunger *et al*, 1998) for refinement. Electron density maps were calculated in CNS and converted for viewing in O (Jones *et al*, 1991) with MAPMAN (Kleywegt, unpublished program). The high R factor and R_{free} are a function of the low quality data used in this study. To obtain better statistics, crystals of higher quality yielding better data will be required. However, as this model was able to answer the questions posed in the literature regarding this ligand, the data were judged of sufficient quality for the time being. The progress of refinement is summarised in Table 4.4.

Table 4.4. Progress of refinement for the 2.7Å model of the con – Man5 complex.

Refinement	**R factor / R_{free}**	
Starting	33.7%	34.4%
CNS : Rigid body fitting of tetramer (30 cycles)	30.5%	30.5%
Rigid body fitting of individual monomers	30.2%	30.4%
O : Four sugar residues of the ligand added.		
CNS : Thermal parameter (30 cycles)		
Positional (60 cycles)	25.1%	28.4%
O : Water molecules added		
CNS : Thermal parameter (40 cycles)	24.4%	27.9%

This final model contains a tetramer of 948 protein atoms (4 subunits of 237 amino acids), 4 Ca^{2+} ions and 4 Mn^{2+} ions, 118 water molecules and 280 sugar atoms. The quality of the data for the final model to 2.7Å is summarised in Table 4.5 and the Ramachandran plot (Ramakrishnan and Ramachandran, 1965), calculated using PROCHECK (Laskowski *et al*, 1993) is shown in Figure 4.3. Glycine residues (not restricted to any particular region of the plot) are represented by a ▲ and non-glycine residues by a ■. The darker the shaded region, the more favourable is the φ, ψ combination. No residues lie outside the additionally

allowed regions. Regions are labelled as follows: A = core alpha, a = allowed alpha, ~a = generous alpha, B = core beta, b = allowed beta ~b = generous beta, L = core left handed alpha, l = allowed left handed alpha, ~l = generous left handed alpha, p = allowed epsilon, ~p = generous epsilon.

Figure 4.3: Ramachandran plot for the 2.7Å con A – ligand model.

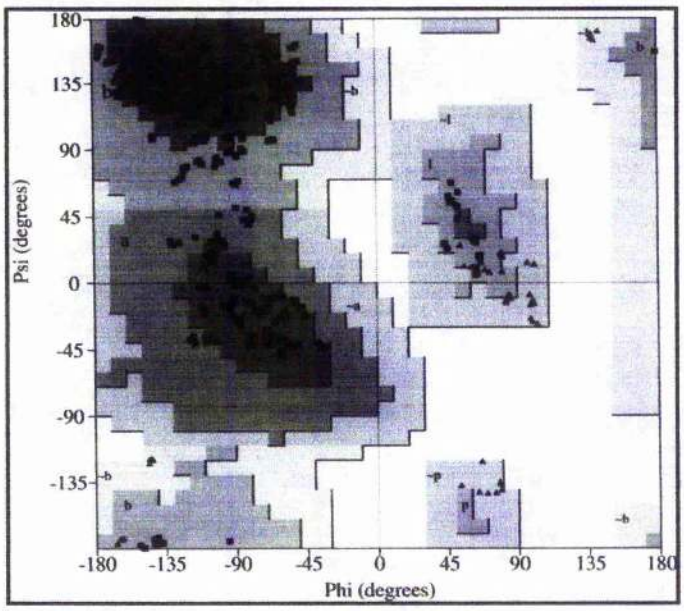


Table 4.5: Quality of the final model for the 2.7Å model of the con A pentasaccharide complex.

Data Collection		
Unique reflections		26890
Completeness of data(%)	(25.00 – 2.7Å / 2.8 – 2.7Å)	84.4 / 82.4
R_{merge} (I) (%)	(25.00 – 2.7Å / 2.8 – 2.7Å)	15.3 / 52.1
Average data redundancy	(25.00 – 2.7Å / 2.8 – 2.7Å)	1.8 / 1.7
% of data > 1σ	(25.00 – 2.7Å / 2.8 – 2.7Å)	78.4 / 52.7
Refinement		
Resolution range (Å)		25.0 – 2.70
R_{free} (%)		27.9
R factor (%)		24.4
Bond r.m.s. deviation[a]		0.008Å
Angle r.m.s. deviation[a]		1.61°
B-factor bounded atoms r.m.s deviation[c]		1.55Å^2
Average NCS r.m.s. difference		0.1
Ramachandran core/additional (%)[b]		81.2 / 18.8
Protein mean B-factor (all)		28.8Å^2

[a] r.m.s. deviation from Engh & Huber ideal values (Engh & Huber, 1991).

[b] Core and additionally allowed regions as defined by PROCHECK (Laskowski *et al*, 1993).

[c] B-factor deviation for bonded atoms

4.4 Discussion

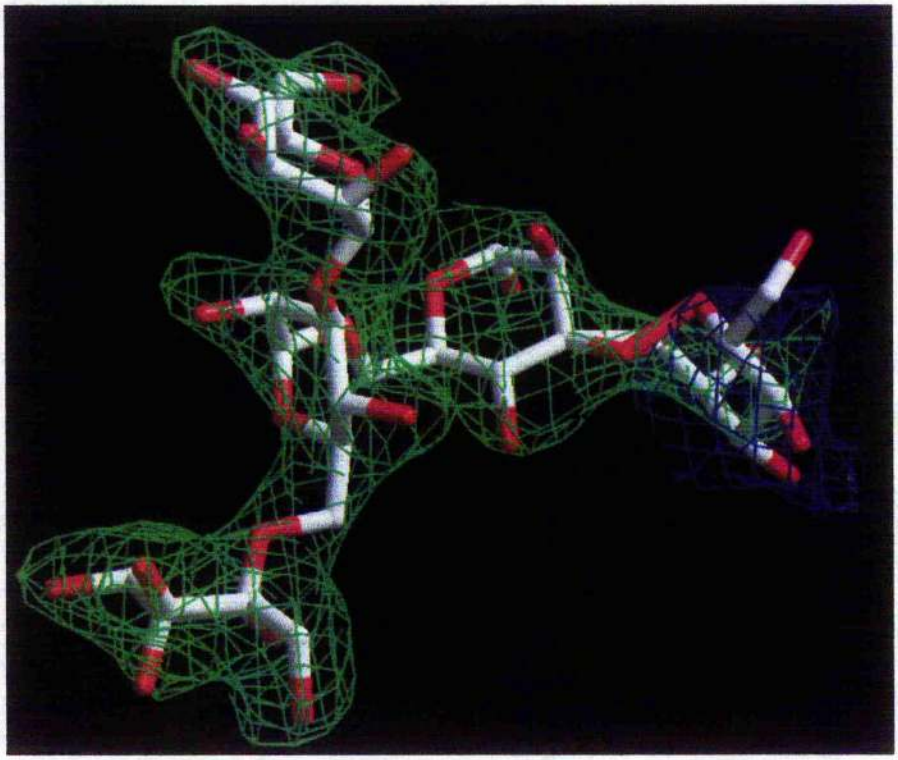
Overall structure of the protein in con A - Man5 complex

In the model of the con A - Man5 complex described here, con A exists as a tetramer where each monomer consists of a sandwich of two β sheets. The overall fold of the protein is identical to the native structure (Deacon *et al.* 1997). The $2F_o$-F_c map is unbroken for the backbone of the structure at 1σ, excluding the commonly found disordered residues at 118-123, 161-163 and 203-205. The unit cell is almost identical to that of the trimannose con A complex (Naismith and Field, 1996). A comparison of the pentamannose structure reported here with the trimannose model shows an average root mean square deviation for all C_α atoms of 0.22Å.

Structure of the bound pentamannose in the con A – Man5 complex.

The pentamannose con A complex displays clear difference density for 4 of the 5 sugar residues in each monomer. For the 5th sugar, the terminal α(1,3) mannose of the branch, convincing density exists in 2 of the 4 subunits, Figure 4.4. Higher quality data may result in all five sugar residues being visualised. In contrast, the previously reported pentamannose - GNA complex (Wright and Hester, 1996), only the trimannose was evident and the branched sugar residues were disordered. GNA binds the trimannose more weakly than succinyl con A, K_a = 3.0 x 10^3M^{-1} compared to 3.4 x 10^5M^{-1} for succinyl con A trimannose binding (Chervenak and Toone, 1995). The lack of order in the 4th and 5th sugar residues of the ligand reflects the non-specific interactions evident in the GNA Man5 complex.

Figure 4.4: Difference density $2F_o$-F_c map (green) contoured at 1σ. The blue density is from an F_o –F_c map contoured at 1σ and corresponds to the position of the 5th the sugar residue.



Conformation of bound pentamannose

Recognition of the pentamannose is essentially at the extended trimannose binding site first described by Naismith and Field (1996), utilising the trimannose moiety present in the ligand. The remaining sugars branch away from the reducing mannose and arc back towards the protein in a manner not unlike a scorpions tail, Figure 4.5. If the φ and ψ angles are as defined in Figure 4.6, then the orientation of the 4th and 5th mannose residues in the crystallographic model is in agreement with previous solution NMR data which predicted a ψ = 60° rotation of the α (1,6) linkage (Wooten *et al*, 1990).

Figure 4.5: Stereo view of the branched sugars arcing back towards the protein making a stabilising H-bond with the main chain via a bridging water molecule. The mannose residues are labelled 1 – 5.

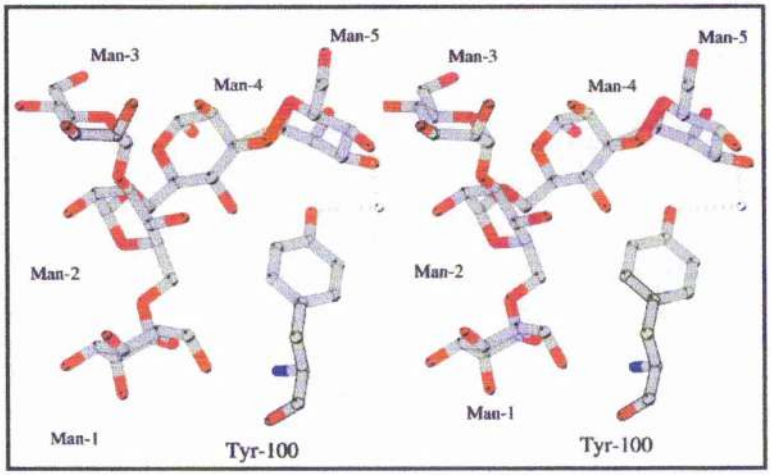
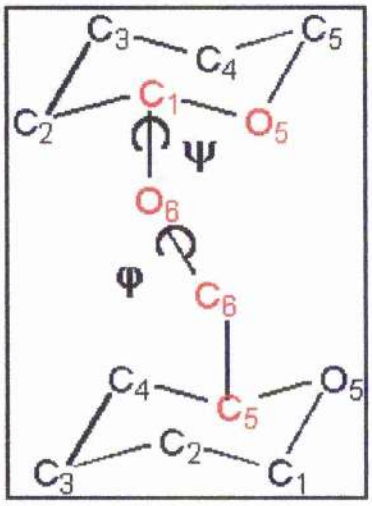


Figure 4.6: The α (1,6) linkage between mannose residues 2 and 4 in Man5. The inter sugar glycosidic torsion angles are defined as $\varphi = C_5\,C_6\,O_6\,C_{1x}$ *and* $\psi = C_6\,O_6\,C_{1x}\,O_5$.



Specific protein – carbohydrate contacts

Only the recognised 'trimannose' of the pentamannose makes direct contact with the protein. It does so through a combination of hydrogen bonds and Van der Waals interactions, Tables 4.4 and 4.5.

As the pentamannose is anchored to the protein through the trimannose moiety of the molecule at the extended trimannose binding site it is not surprising to find the interactions between the saccharide and con A are primarily those discussed by Naismith and Field (1996).

Table 4.6: Hydrogen bonding and polar contacts less than 3.5Å between the con A and the pentamannose.

Sugar	**Protein**	**Distance Å**
1, 3 terminal mannose (3)		
O-3	Thr-15 OG1	2.91
O-3	Thr-15 N	2.73
O-3	Pro-13 O	2.78
O-4	Thr-15 OG1	2.51
O-4	Asp-16 N	2.96
O-4	Asp-16 OD1	3.07
reducing mannose (2)		
O-2	OW	2.75
O-4	Tr-12 OH	2.46
1,6 terminal mannose (1)		
O-3	Arg-228 N	2.93
O-4	Asn-14 ND2	3.16
O-4	Asp-208 OD2	2.47
O-5	Leu-99 N	2.84
O-6	Tyr-100 O	3.14
O-6	Tyr-100 N	2.79
O-6	Leu-99 N	3.01
1,3 Branched Mannose (5)		
O-4	OW	2.70

Table 4.7: Van der Waals contacts less than 4.0Å

Sugar	Residue
1, 3 terminal Mannose (3)	Tyr-12, Pro-13, Asn-14, Thr-15, Asp-16
Reducing Mannose (2)	Tyr-12, Asp-16, Leu-99, Arg-228
1, 6 Terminal Mannose (1)	Tyr-12, Asn-14. Gly-98, Leu-99, Tyr-100 Asp-208, Gly-227, Arg-228
1, 6 Branch Mannose (4)	Leu-99
1, 3 Branched Mannose (5)	Tyr-100

The α(1,6) Mannose α(1,3) Mannose branch (sugars 4 and 5)

As noted above neither of the 4th or 5th mannose residues form any direct interaction with the protein. Only one water molecule acts to stabilise these sugars through hydrogen bonding. O4 of the terminal sugar in the branch makes contact with the protein (Tyr-100 OH) via a bridging water molecule, Figure 4.5. The non-displacement of the trimannose core from the extended binding site coupled with the non-specific recognition of the 4th and 5th sugar residues reflects the similar affinity displayed by con A for the trimannose and the pentamannose discussed in this chapter. The additional interactions observed with mannoses 4 and 5 of the ligand help stabilise the overall ligand conformation. In addition, as the pentamannose is recognised by con A as the 'trimannose' then any crosslinking bridge could only involve the 4th and 5th sugars of the oligosaccharide, that is a maximum of around 11 C - C bond lengths. Molecular modelling has shown that this is too short to allow the tethering

of two recognition domains on separate con A molecules and cross-linking would be prevented by steric interactions.

4.5 Conclusions

It is clear from this structure that the recognition of higher order oligosaccharides to con A is via the trimannoside moiety of the molecules. That this pentamannose does not crosslink con A is probably an accumulation of factors. Firstly, before any aggregation could take place the trimannose core would have to be displaced from the binding site. The rate of aggregation would then be controlled by the rate at which the trimannose re-bound. That is, aggregation would be kinetically controlled between a trade off of the displacement the core and subsequent aggregation of subunits, against the re-binding of the trimannose core and the complex existing as a single entity. In addition, the size of the non-specifically recognised branch of the ligand (4^{th} and 5^{th} sugars) appears to be too short to allow the tethering of adjacent con A subunits and would incur severe steric penalties in the absence of the displacement of the trimannose core. As sugars 4 and 5 of the ligand adopt a conformation which allows only non-specific interactions with the protein reflects the almost identical thermodynamic properties of this ligand to those displayed by the trimannose and highlights errors in the literature where it was anticipated this oligosaccharide would crosslink con A monomers.

4.6 Future Work

The first improvements to this model should involve the collection of better data, this will ultimately require better crystals of the complex. To probe the binding of these higher order oligosaccharides further still, a three-dimensional model of a larger branched oligosaccharide would be desirable. Either of the Man 7, 8 or 9 oligosaccharides discussed in the introduction of this chapter which apparently display upto 7-fold greater affinities for con A would appear more likely to crosslink this lectin.

Chapter 5

Initiating A Crystallographic Study of 3-Methylaspartate Ammonia Lyase

5.1 Summary

Methylaspartate ammonia lyase (MAL) from *Clostridium tetanomorphum* has been overexpressed and purified in two different cell lines. Initially the enzyme was expressed in the constitutive strain pSG4 which required some 16 hours incubation. By a process of anion exchange and hydrophobic interaction chromatography the enzyme was purified to 95% homogeneity, judged by SDS and IEF PAGE. Attempts to purify the enzyme further have been unsatisfactory, however dynamic light scattering experiments confirmed the enzyme to be monodisperse. Purification yields less than 1 mgL^{-1} culture. Small crystals (0.1 x 0.1 x 0.1mm) have been obtained in hanging drops when the enzyme is equilibrated against $(NH_4)_2\ SO_4$. The crystals only diffract to 6Å and attempts to acquire an accurate unit cell were hampered by this.

MAL has also been overexpressed in two high efficiency constructs. Overexpression in pET16b BL21 (DE3) yields some 8.1 mgL^{-1}. As the enzyme expressed from this system contains 6 consecutive histidine residues at the N-terminus it can be purified to near 100% homogeneity via nickel chelation and anion exchange chromatography. However, light scattering experiments showed the enzyme to be polydisperse and subsequently crystallisation of the enzyme failed. Attempts to selectively cleave the tag proved unsatisfactory.

Similarly, overexpression in pET11a BL21 (DE3) yields 7.1 mgL^{-1} culture after purification. Despite these results this enzyme also failed to crystallise. Sequencing of the clone revealed a mutation relative to the constitutive strain, resulting in a C-terminal extension. It appears this extension is not conducive to crystallisation.

5.2 Introduction

It is common that specialty chemicals synthesised by pharmaceutical companies require long and complex synthetic pathways. On an industrial scale these processes often become highly expensive, generate pollution and result in low yields. This problem is particularly acute for bioactive compounds which are often only active in one enantiomeric form, and in the worst case scenario the wrong enantiomer may have unwanted side effects, for example Thalidomide.

Enzymes, by their nature, are highly stereo and regio selective and are often extremely efficient catalysts. As enzymes work in water they do not require organic solvents or elevated temperatures, eliminating the expensive and pollutive disposal measures which organic waste requires. If it were possible to harness the 'natural skills' of biochemical enzymes within industrial processes this would go some way to alleviating the problems of pollution and expense. Essentially, biotransformations hold out the promise of new efficient and highly selective synthetic routes.

One example where x-ray crystallography has aided the development of a new series of enzymes is in the work described by Withers (1999) on glycosidases. The determination of the three-dimensional structures of several glycosyl-enzyme complexes has identified the amino acids directly involved in catalysis. As a result a new class of enzymes have been developed where glycosidic bonds are selectively synthesised but not hydrolysed allowing the controlled synthesis of oligosaccharides upto 5 sugars in length with yields upto 90% in individual steps.

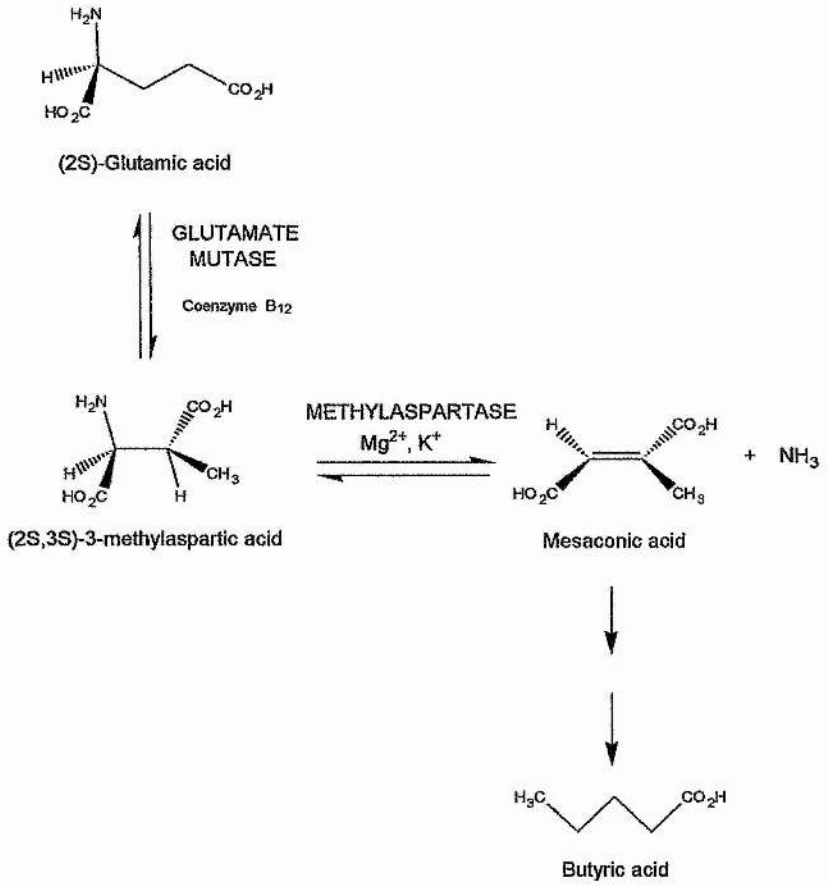
α, β - unsaturated amino acid residues play important roles in many biochemical reactions allowing additional new functionality to be added whilst retaining stereo control of the

reaction. Conventional methods of achieving this reaction are difficult and expensive on a commercial scale.

Methylaspartate ammonia lyase

MAL occurs in the obligate anaerobe *Clostridium tetanomorphum,* where it functions in the main metabolic pathway for glutamate (Barker *et al,* 1958). The enzyme catalyses the regio and stereoselective addition, and the reverse elimination, of ammonia from (2S, 3S)-3-methylaspartic acid yielding mesaconic acid which is further processed to glutamic acid Figure 5.1. The enzyme is a member of the ammonia lyase family of proteins, for which no X-ray structure currently exists. The *Clostridial* enzyme is a 91 000 Da homodimer of 413 amino acids where both monovalent and divalent cations are required for optimal enzyme activity; potassium and magnesium having been shown to be most effective (Goda *et al,* 1992). The enzyme can process various analogues of the natural substrate including the L-*erythro*-(2S, 3R)- diastereoisomer of methylaspartic acid (Barker *et al*, 1958; Archer and Gani, 1993) as well as (2S)-aspartic acid and a number of three 3-alkyl homologues (Winkler and Williams, 1967). In addition the enzyme is capable of providing access to various 3-substituted aspartic acids by direct amination of the substituted fumaric acids, and has been shown to tolerate some non-ammonia N-nucleophiles (Akhtar *et al,* 1987). The enzymes ability to accommodate different substrates and ammonia surrogates reveals its potential use in the synthesis of homo chiral α-substituted β-amino acids.

Figure 5.1: The methylaspartase reaction.



In the other ammonia–lyase enzymes a dehydroalanine residue has been implicated in the mechanistics of the processes involved and has been suggested to perform a similar function for methylaspartase (Goda *et al*, 1992). The source of the dehydroalanine in both phenylalanine and histidine ammonia lyases is the serine residue of the catalytic triad Ser–Gly–Asp. This sequence is common to both histidine and phenylalanine ammonia lyases and is not surprisingly also found in the sequence of MAL (Goda *et al*, 1992). The mechanism by which methylaspartase operates has been probed in recent years by Gani and

co–workers (Goda *et al*, 1992; Botting and Gani, 1992; Pollard *et al*, 1998). Most recently they postulate that each true substrate forms a covalent adduct with the enzyme through the amination of a dehydroalanine residue (Goda *et al*, 1992). Substrates are N–protonated and the alkylammonium groups are deprotonated internally in a closed solvent-excluded pocket after a K^+ ion has bound (Botting and Gani. 1992). The enzyme-substrate complex is subsequently deaminated in two elimination processes (Goda *et al,* 1992). (2S, 3S)-3-methylaspartic acid, the natural substrate for the enzyme, is eliminated in a concerted process (Botting *et al*, 1989) as is (2S)-aspartic acid. However, the L-*erythro*-(2S, 3R)-diastereoisomer of methylaspartic acid appears to be eliminated via a carbocationic mechanism, suggesting that a change in stereo chemistry at C-3 induces a change in mechanism (Pollard *et al*, 1998).

The ultimate aim of this research is to redesign the active site of methylaspartase and determine a structural basis for the mechanism of the reaction. This will be achieved through a combination of site directed mutagenesis, molecular modelling and structure determination. Mutants of the enzyme and subsequent crystal structures of the mutant protein along with protein-inhibitor complexes would permit characterisation of the enzyme mechanism, which may be unprecedented in biology.

Assay

The enzyme was assayed according to the method of Barker (Barker, 1958) where one unit of enzyme is defined as the amount that catalyses the formation of 1μmol of mesaconic acid at pH 9.0 and 30°C. This is determined by the increase in OD_{240} in the presence of 4mM (2S,3S)-3-methylaspartic acid, 50mM TRIS-HCl, 10mM KCl, and 1mM $MgCl_2$.

5.3 Experimental

Constitutive Expression System

All materials used were purchased from Sigma (Poole, UK) unless otherwise stated. Prior to all cultivation procedures, all containers, pipette tips and media were sterilised by autoclaving at 120°C for 30 minutes.

Overexpression of E. coli pSG4

A single colony of *E. coli* pSG4 was selected from an ampicillin (200μgmL^{-1}) agar plate and used to inoculate 5 mL of LB media in the presence of 200μgmL^{-1} ampicillin. This culture was incubated at 37°C, 90rpm and the A_{600} monitored until it reached 0.6. At this point the cultures were used to inoculate 500ml of similarly treated media. This culture was incubated for a further 16 hours at 37°C, 90rpm. The cells were harvested by centrifugation (10000*g*, 4°C, 30 minutes) and frozen at –78°C until required.

Preparation of Cell Extracts and Ammonium Sulphate Fractionation

Frozen cell paste was thawed in ice-cold lysis buffer (50mM TRIS pH 7.6, 1mM DTT, 1mM PMSF) at a concentration of 0.5g wet cell paste per mL buffer and the cell contents released by sonication (140kW, 20kHz) at 4°C for 6 x 30 seconds. Cell debris was pelleted by centrifugation (20000*g*, 4°C, 30 minutes) and the resulting supernatant brought to 40% $(NH_4)_2SO_4$ saturation by the slow addition of 24.2g $(NH_4)_2SO_4$ per 100mL^{-1} at 4°C. This solution was left to stir for 30 minutes at 4°C. After centrifugation (20000*g*, 4°C, 10 minutes) the supernatant was brought to 80% $(NH_4)_2SO_4$ saturation by slow addition of a further aliquot of $(NH_4)_2SO_4$. This solution was left to stir for a further 30 minutes at 4°C

before centrifugation (20000*g*, 4°C, 10 minutes). The pellet obtained contained the target enzyme, MAL, with assays confirming ~90 % enzyme recovery. The resultant pellet was resuspended in lysis buffer and dialysed for 3 x 40 minutes against 20mM TRIS, pH 7.5.

Purification and Characterisation of MAL

100mg of total protein were 0.02μm filtered before being applied to an anion exchange chromatography column (Perseptive Biosystems, HQ) pre–equilibrated in buffer A (20mM TRIS, pH 7.5). Elution was over an increasing gradient from 0 to 500mM NaCl in buffer A. Initially all peaks were collected and assayed for activity after which only the active peak was pooled and concentrated by ultrafiltration. Dialysis (3 x 40minutes) against 50mM KPi, 40% $(NH_4)_2SO_4$, pH8.5 followed before application to a hydrophobic interaction chromatography column (Perseptive Biosystems, HP). 50mg of total protein was applied to the column pre–equilibrated with 50mM KPi, 40% $(NH_4)_2SO_4$, pH8.5. Elution was achieved over a decreasing $(NH_4)_2SO_4$ gradient (40 to 0%). Again all peaks were initially pooled to identify the active fraction, after which only the methylaspartase peak was collected. Enzyme was concentrated by ultrafiltration ahead of dialysis (3 x 40 minutes, 20mM TRIS, 10mM KCl, 1mM $MgCl_2$, 1mM DTT) in preparation for crystallisation. Analysing enzyme purity by both SDS and IEF PAGE showed the enzyme to be greater than 90% homogenous. Dynamic light scattering experiments revealed a slightly polydisperse solution, Table 5.1. The average polydispersity value as a function of the hydrodynamic radius was 23%. Knowing that for successful crystallisation monodisperse solutions are desired, attempts were made to further purify the enzyme. Throughout purification enzyme activity was

monitored spectrophotometrically and concentration by the Bradford method (Bradford, 1976), Table 5.2.

Table 5.1: Dynamic light scattering results for constitutively expressed MAL. The average polydispersity value as a function of the hydrodynamic radius is 23 %.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDA	Temp. ^{0}C	Count Rate	Base Line	% Radius / PolyD
1	0.713	621	3.6	0.975	66	20.2	180	1.003	27
2	0.725	608	3.6	1.048	62	20.5	183	1.002	29
3	0.714	615	3.6	1.069	65	20.4	177	1.002	29
4	0.725	621	3.6	0.834	68	20.0	183	1.001	23
5	0.729	605	3.7	0.816	66	20.3	183	1.001	22
6	0.725	614	3.6	0.968	64	20.5	184	1.000	26
7	0.729	601	3.6	1.063	62	20.5	179	1.002	29
8	0.735	605	3.6	0.826	65	20.4	180	1.001	23
9	0.736	615	3.6	0.822	69	20.4	181	1.000	22

Table 5.2. Yield of constitutively expressed protein obtained after subsequent purification steps.

	Vol (mL)	Protein (mg)	Total Activity (U)	*Specific Activity* (U/mg)	Yield (%)
Crude Extract	24	201	2500	12.4	100
$(NH_4)_2SO_4$	3	97	2180	22.5	87
Post HQ	3	10	1440	144	60
Post HP	1	1	870	290	35

It was previously noted that at pH values below 7 no enzyme activity is recorded. This eliminated cation exchange chromatography as an additional purification method. Size exclusion chromatography appeared the only viable alternative. Three different gel

permeation columns (Sephadex G – 150, Superose 7 and Sepharose 75) were tried in order to achieve some form of improvement in enzyme purity, however analysing protein purity afterwards by electrophoresis and light scattering displayed no obvious enhancement. The yield of protein obtained from this expression system (~ 1mg L^{-1}) was not conducive for crystallisation studies. An overexpression system, where a higher percentage of the soluble protein expressed was the target protein, would likely result in a higher yield of more pure enzyme.

Crystallisation of Constitutively Expressed MAL

In spite of a less than ideal level of enzyme homogeneity, crystallisation of the enzyme employing the sparse matrix approach utilising the vapour diffusion methodology was attempted (Ducruix and Giege, 1992). Commercially available Crystal Screen Kits I and II from Hampton Research were used for initial screens along with $(NH_4)_2SO_4$, NaCl, PEG, PEG / LiCl and MPD grid screens. After 6 weeks incubation at 20.5°C small crystals of dimensions 0.1 x 0.1 x 0.1mm were observed with 1.8M $(NH_4)_2SO_4$, 0.1M HEPES pH 7, Table 5.3.

A small crystal was mounted in a thin walled glass capillary and sealed in the presence of the precipitating agent. X–rays were generated using a FR–951 rotating anode (Enraf Nonius) with copper target and focussed with the MacScience mirror system. With operating parameters at 50kV, 100mA, λ = 1.54Å, 1° oscillation, exposure time 60 minutes and a crystal to detector distance of 150mm, data were observable to 6 Å as recorded on the DIP 2000 image plate. The unit cell could not be determined with any certainty from this data, though preliminary analysis suggested it to be rhombohedral.

Table 5.3: Initial screens performed in crystallisation of MAL.

Precipitant	pH	Protein Concentration mg ml^{-1}	Mixing	Crystals
PEG 6K	4 - 9	10	Yes	No
PEG 6K / Li Cl				No
$(NH_4)_2SO_4$				Yes
MPD				No
Na Cl				No
Hampton CS I				No
Hampton CS II				No

The cloning of MAL into a pET vector offers many advantages from the previous expression system. Overexpression of the enzyme, in a pET vector, is controlled by the addition of the artificial inducer isopropyl-β-D-thiogalactopyranoside (IPTG). However, the host must contain a chromosomal copy of the gene for T7 RNA polymerase. Such hosts are lysogens of bacteriophages DE3 which carry the *lacUV5* promoter and the gene for T7 RNA polymerase (Studier and Moffat, 1986). Once the lysogen is formed, the only promoter known to direct transcription of the T7 polymerase is *lac UV5,* which is inducible with IPTG. Consequently, addition of IPTG to a growing culture of the lysogen induces T7 RNA polymerase, and the target DNA is transcribed in the plasmid. As T7 RNA polymerase is very selective, almost all the cells energies are tuned to target gene expression. As a result of this as much as 50% of the total cell protein may be the desired product. In addition, a poly-histidine tag (6 consecutive histidines at the N-terminus of the protein, HIStag) was included in this expression system. This allows near 100% purification of the enzyme to be achieved in one step utilising the tags specific affinity for nickel (Janknecht *et al*, 1991).

Overexpression of E. coli pET16b BL21 (DE3) MAL

A single colony of *E. coli* pET16b BL21 (DE3) containing the poly his-tagged gene for MAL was selected from an ampicillin ($200\mu gmL^{-1}$) agar plate and used to inoculate 5 mL of LB media in the presence of $200\mu gmL^{-1}$ ampicillin. This culture was incubated at 37°C, 150rpm and grown to an optimum cell density (OD_{600} 0.6) before the addition of 1mM IPTG. Incubation continued for a further 3 hours under the same conditions. The cells were subsequently harvested by centrifugation (10000*g*, 4°C, 30 minutes) and frozen at –78°C until required.

Preparation of Cell Extracts

Frozen cell paste was thawed in ice-cold lysis buffer (50mM KPi, pH 8.0, 1mM DTT, 1mM PMSF) at a concentration of 0.5g wet cell paste per mL buffer before the cell contents were released by sonication (140kW, 20kHz) at 4°C for 6 x 30 seconds, resting on ice in between. Cell debris was pelleted by centrifugation (20000*g*, 4°C, 30 minutes).

Purification of MAL by Ni Chelation and Anion Exchange Chromatography.

The cell extract was 0.02μm filtered before being loaded to a Ni-NTA Superflow affinity column (Qiagen) previously equilibrated in lysis buffer poisoned with 50mM imidazole. Column washing continued with equilibration buffer until the OD_{280} of column wash matched that of the equilibration buffer. MAL was eluted in an increasing lysis buffer / imidazole gradient (50 to 500mM). All active fractions were pooled and concentrated by ultrafiltration before dialysis against 20mM TRIS, pH 7.5. 50mg total protein were 0.02μm filtered before being applied to an anion exchange chromatography column (Perseptive Biosystems, HQ) pre–equilibrated in 20mM TRIS, pH 7.5. Elution was over an increasing

salt gradient from 0 to 500mM NaCl. Active peaks were pooled and concentrated by ultrafiltration to around 10 mg ml^{-1} and dialysed against 20mM TRIS pH7.5, 10mM KCl, 1mM $MgCl_2$, 1mM DTT, 1mM EDTA in preparation for crystallisation. Throughout purification enzyme activity was monitored spectrophotometrically and content by the Bradford assay, Table 5.4.

Characterisation of E. coli pET16b BL21 (DE3) MAL.

Dynamic light scattering experiments were performed on uncleaved enzyme at a range of concentrations (2 – 15 mg ml^{-1}) and results were indicative of a polydisperse suspension, Table 5.5. The average polydispersity value as a function of the hydrodynamic radius was greater than 35 % in all instances. However, the protein did migrate as a single band on SDS - PAGE, Figure 5.2.

Cleavage of the HIStag by Factor Xa

In an effort to selectively cleave the HIStag tail, Factor Xa was added to aliquots of enzyme at concentrations of 0.6, 1.2 and 1.6 units/mg total protein and incubated at room temperature for 6, 12 and 18 hours. A specific cleavage site had been inserted in the gene for this purpose. However, cleavage was not specific. Visualisation of the products on SDS gels revealed many products after incubation of the enzyme with Factor Xa, Figure 5.3.

Crystallisation of E. coli pET16b BL21 (DE3) MAL

Crystallisation trials were undertaken employing the sparse matrix approach utilising the vapour diffusion methodology. Commercially available Crystal Screen Kits I and II from

Hampton Research were used for initial screens along with $(NH_4)_2$ SO_4, NaCl, PEG, PEG / LiCl and MPD grid screens.

After 52 weeks incubation at 20.5°C with periodic viewing no crystalline material was observed.

With the failure of the HIStag variant of the enzyme to crystallise, together with the inability to selectively cleave the tag, it was concluded a non-tagged version of the enzyme was required.

Table 5.4. Yield of E. coli pET16b BL21 (DE3) MAL protein

	Vol (mL)	Protein (mg)	Total Activity (U)	*Specific Activity* (U/mg)	Yield (%)
Crude Extract	30	412.5	12840	29	100
$(NH_4)_2SO_4$	16	54.0	10656	197	83
Post HQ	20	40.5	7792	192	60
Post HP	2	8.1	4675	192	36

Table 5.5: Dynamic light scattering results for E. coli pET16b BL21 (DE3) MAL. The average polydispersity value as a function of the hydrodynamic radius is 43 %.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDA	Temp. °C	Count Rate	Base Line	% Radius / PolyD
1	0.853	184	12.5	4.895	1331	23.2	438	1.006	41
2	0.858	193	12.6	4.952	1368	23.5	423	1.006	39
3	0.848	180	12.8	6.341	1423	23.4	421	1.006	49
4	0.849	180	12.9	5.044	1430	23.5	420	1.005	39
5	0.843	181	13.0	6.398	1455	23.9	434	1.008	49
6	0.85	175	13.2	6.515	1521	23.5	432	1.006	49
7	0.850	179	13.0	5.094	1465	23.5	421	1.005	49
8	0.851	180	12.9	6.259	1445	23.4	414	1.005	39
9	0.865	181	12.9	4.787	1419	23.4	413	1.008	39

Figure 5.2: Silver stained gel displaying MAL.

Lane1: Molecular markers (kDa)
Lane2: Post $(NH_4)_2 SO_4$ fractionation.
Lane3: Post anion exchange chromatography.
Lane4 & 5: Post hydrophobic interaction chromatography.

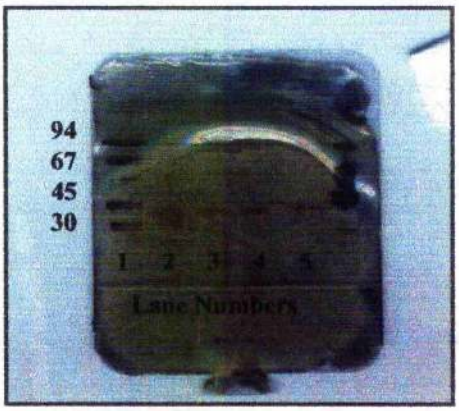


Figure 5.3: Silver stained gel displaying attempts to cleave the HIStag by factor Xa.

Lane1 & 11: Untreated homogenous MAL.
Lane2: 60 minutes incubation 0.6 units factor Xa.
Lane3: 120 minutes incubation 0.6 units factor Xa.
Lane4: 30 minutes incubation 1.2 units factor Xa.
Lane5: 60 minutes incubation 1.2 units factor Xa.
Lane6: 120 minutes incubation 1.2 units factor Xa.
Lane7: 30 minutes incubation 1.6 units factor Xa.
Lane8: 60 minutes incubation 1.6 units factor Xa.
Lane9: 120 minutes incubation 1.6 units factor Xa.
Lane10: Overnight incubation 1.2 units factor Xa

Cloning of MAL in pET 11a

Tracey Neal performed the cloning of the new construct in David Gani's laboratory.

Expression of Recombinant MAL in E. coli pet11a BL21 (DE3)

5 mL of terrific broth medium containing ampicillin was inoculated with a single colony and incubated overnight at 37°C. This culture was used to inoculate 500ml of similarly treated terrific broth and the resulting solution was shaken at 150rpm at 37°C until an A_{600} value of 0.6 – 0.8 was reached. Protein production was induced by the addition of IPTG to a final concentration of 1mM and incubation continued for 3 hours. Cells were harvested by centrifugation (10000*g*, 4°C, 30 minutes) and stored at –78°C until required.

Preparation of Cell Extracts

Cells from cultures of pet11a BL21 (DE3) were resuspended in ice-cold lysis buffer (50mM KPi, pH 8.0, 1mM DTT, 1mM PMSF) at a concentration of 0.5g of cells per mL buffer. Suspended cells were sonicated on ice for 6 x 30 seconds with a cooling period of 30 seconds in between. Cell debris was pelleted by centrifugation (20000g, 4°C, 30 minutes) and the cell supernatant transferred to a clean tube and used in cell-free assays.

Purification of Recombinant E. coli pET11a BL21 (DE3) MAL.

Crude extract was brought to 20 % $(NH_4)_2SO_4$ concentration by the slow addition of 10.7g $(NH_4)_2SO_4$ per 100 ml of extract. Precipitated protein was pelleted by centrifugation (20000*g*, 4°C, 30 minutes) and the supernatant transferred to a clean tube. $(NH_4)_2SO_4$ was

removed by dialysis against buffer A (20mM TRIS, pH 7.5, 2mM DTT) in preparation for anion exchange chromatography.

A 7.84ml anion exchange column (Perseptive Biosystems, HQ) was equilibrated in buffer A and 100mg's of total protein was applied at a flow rate of 5ml min^{-1}. After column washing (3 column volumes of buffer A), protein was eluted with a gradient of 0 to 500mM NaCl in buffer A. Protein was detected on-line at 280nm and fractions were assayed for biological activity. Active fractions were pooled from successive runs and concentrated by ultrafiltration.

Protein from anion exchange was dialysed for three changes against buffer B (50mM KPi, 40% $(NH_4)_2SO_4$, pH8.5) before application to a hydrophobic interaction chromatography column (Perseptive Biosystems, HP) pre-equilibrated in buffer B. After washing the column in 3 volumes of buffer B protein was eluted by application of a 40 - 0% $(NH_4)_2SO_4$ gradient over 15 column volumes. Again protein was detected on-line at 280nm and active fractions detected by assay. Protein content was measured by Bradford, Table 5.6, and analysed by SDS-PAGE.

Table 5.6. Yield of E. coli pET11a BL21 (DE3) MAL protein obtained after subsequent purification steps.

	Vol (mL)	Protein (mg)	Total Activity (U)	Specific Activity (U/mg)	Yield (%)
Crude Extract	160	3888	49040	12.6	100
$(NH_4)_2SO_4$	19	548	43510	79.4	89
Post HQ	18	117	31000	265	63
Post HP	18	71	23606	332.5	48

Characterisation of Recombinant E. coli pET11a BL21 (DE3) MAL.

Dynamic light scattering experiments were performed on purified enzyme. The average polydispersity value as a function of the hydrodynamic radius was 39%.

Crystallisation of Recombinant E. coli pET11a BL21 (DE3) MAL.

$(NH_4)_2SO_4$ was removed by dialysis against buffer A in preparation for crystallisation.

As before the crystal screens from Hampton Research were used to screen for initial crystallisation conditions along with the following grid screens; PEG 6K, PEG 6K / LiCl, $(NH_4)_2SO_4$, MPD and NaCl. All experiments were performed as hanging drops at a range of protein concentrations, Table 5.7.

Table 5.7: Screens performed in crystallisation of E. coli pET11a BL21 (DE3) MAL

Precipitant	**pH**	**Protein Concentration mg ml^{-1}**	**Mixing**	**Crystals**
PEG 6K	4 - 9	5, 10 & 15	Yes	No
PEG 6K / Li Cl				No
$(NH_4)_2SO_4$				No
MPD				No
NaCl				No
Hampton CS I				No
Hampton CS II				No

As with the previous high efficiency expression strain of the enzyme no crystals were apparent after several weeks' incubation at 20.5°C. As the enzyme was still biologically active, the failure to reproduce the initial crystallisation from the constitutive strain with a cleaner more active enzyme was puzzling.

Sequence Characterisation

Automated sequencing of pET11a BL21 (DE3) MAL was performed using dye deoxy methodology and one mutation was uncovered at the C-terminus of the protein. The 'true' **TAA** stop codon had mutated to **TAT**, encoding a tyrosine residue, which allowed for a further 6 amino acids to be transcribed, before another **TAA** stop codon was encountered. The extension included two glycine and one cysteine residue.

5.4 Results and Discussion

MAL has been overexpressed and purified from *E. coli* pSG4. However, levels of protein obtained (1mg ml^{-1}) are so low that a crystallographic study of the enzyme would be extremely time consuming. Despite this limiting factor, small crystals of the enzyme have been grown after equilibration against $(NH_4)_2SO_4$ and X–ray data have been recorded in-house to 6 Å. Given that the crystals were only $0.1mm^3$ and suffered from radiation damage, the low resolution diffraction obtained is not surprising.

To try and increase the yield of soluble protein obtained per preparation, the enzyme was overexpressed in a high efficiency vector; *E. coli* pET16b BL21 (DE3). Using this vector system the enzyme was expressed under the control of a T7 *lac* promoter allowing enzyme production to be initiated by the addition an artificial inducer, IPTG. Subsequently, a larger quantity of soluble protein should be obtained. An additional feature of this clone is the presence of an extra 6 consecutive histidine residues at the N terminus which can be selectively cleaved by the enzyme Factor Xa. With the HIStag, purification is simplified by the inclusion of an affinity chromatography step utilising the ability of nickel to selectively bind the tag. Approximately 10 times more soluble protein was produced with this system

and the modified enzyme showed activity on a par with the native (constitutive enzyme). However, although the method of expression and the purification protocol employed resulted in increased amounts of pure enzyme, results from dynamic light scattering were not promising for crystallisation. The initial intention was to cleave the tag before crystallisation, but as it did not interfere with enzyme activity attempts to crystallise the tagged enzyme seemed justified. Crystallisation trials employing over 300 different conditions at various protein concentrations and temperatures failed to grow crystals. These results were not unexpected given the dynamic light scattering results. Repeating these crystallisations after cleavage of the tag was the next logical step but the inability to selectively cleave the tag made this impossible. Factor Xa cleaves after the sequence IEGR. Inspecting the primary sequence of MAL, the sequence IEGP is identified starting at residue 271. It seems likely that this sequence may be acting as a pseudo cleavage site, which could result in the multiple products discovered in practice. Alternatively, small amounts of contaminating thrombin activity are known to be present (Boehringer Mannheim, 1993) in Factor Xa, so cleavage at very exposed thrombin sites cannot be dismissed.

Several experiments were performed with the intention of introducing order into the tag, as it was believed the flexible nature of 6 histidines located at the N terminus might be inhibiting crystallisation. As dynamic light scattering suggested extremely high molecular weights it was presumed extensive cross-linking was taking place. To try and circumvent this problem EDTA (to complex any metals) and detergent (to further solubilise the solution) were added. Neither additive offered a solution with the dynamic light scattering results remaining unchanged.

As an alternative approach to introduce order to the tag the addition of chelating agents was examined. However, the addition of either one of the transition metals nickel and zinc at 1mM induced immediate precipitation. It had been hoped the metals would serve to chelate the tag internally and possibly limit its movement, yet as the solutions precipitated, this suggests that with two free coordination sites (Ni^{2+} and Zn^{2+}), the metals were actually promoting more extensive cross-linking. It appeared the tag designed to aid in purification of the enzyme created more problems than it solved and ultimately prevented crystallisation. For this reason the enzyme was recloned in a tag free high efficiency vector pET11a before overexpression and purification. Again crystallisation proved fruitless. Sequencing of the enzyme revealed the C-terminal extension, including two glycine residues and a cysteine. Giving the nature of the alternation relative to the native enzyme the failure to re-crystallise the enzyme is not surprising.

5.5 Conclusions and Future Work

It is apparent that no variant of the enzyme presented here is favourable to work with given the ultimate aim of the project. The constitutive strain does not express sufficient quantities of enzyme that would allow a comprehensive search for crystallisation conditions. Whereas, although the pET strains overexpress large amounts of soluble protein, the protein has failed to crystallise for a variety of reasons.

It is clear that for this project to succeed, the enzyme will have to be cloned into an expression system that allows large amounts of enzyme to be expressed, but is lacking in any additional tags. There are several commercially available expression systems that will facilitate this process and work is ongoing within David Gani's laboratory to rectify this situation.

Chapter 6

Initiating a Crystallographic Study of α-D-Glucose-1-Phosphate Thymidylyltransferase and TDP-Glucose 4,6-dehydratase

6.1 Summary

RmlA from *Streptococcus suis* has been overexpressed and purified to homogeneity suitable for crystallisation. The enzyme migrates as a single band at 30kDa on denaturing gels and has an average polydispersity of 30.5%. The enzyme has been crystallised from both PEG / LiCl and $(NH_4)_2SO_4$. Currently only crystals from PEG / LiCl have been characterised. The crystals (0.1 x 0.1 x 0.1mm) diffract to around 8Å in-house and 3.5Å at the synchrotron radiation source, Daresbury. Further optimisation of the crystallisation parameters is required.

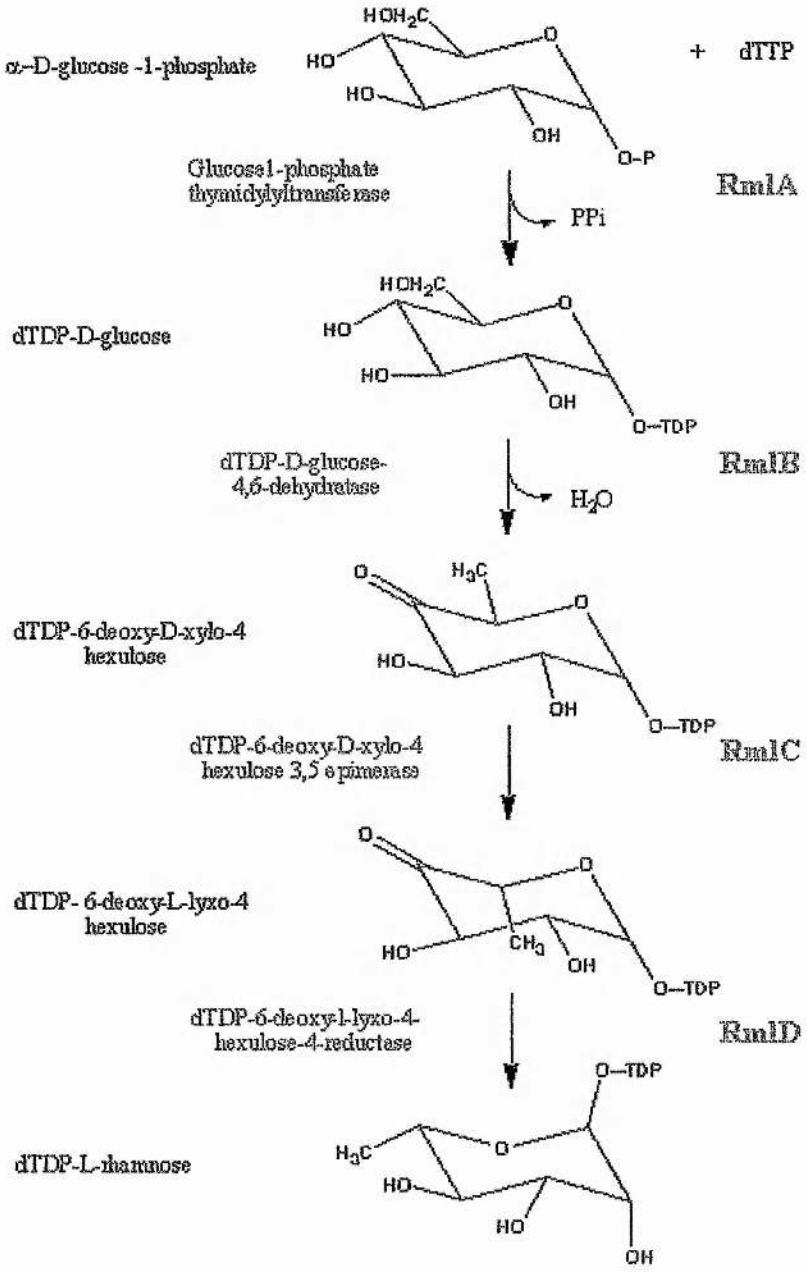
A seleno-methionine variant of RmlB from *Salmonella enterica* has also been overexpressed, purified and crystallised in a form suitable for x-ray data collection. The enzyme migrates as a single band at 44kDa on denaturing gels and crystallises after 14 days equilibration against 0.1M MES pH 6.3, 1.5M $LiSO_4$, 2mM DTT. A partially complete data set (85%) has been collected in-house to 3.1Å.

6.2 Introduction

L – rhamnose (L – rha), like galactofuranose (chapter 2), is a key component of the cell wall of many pathogenic bacteria including mycobacteria (McNeil *et al*, 1990). As in the case of gal_f, its presence is essential for cell wall integrity being the linker between the inner peptidoglycan layer and arabinogalactan Figure 1.6. From this it is clear that the inhibition of L – rha biosynthesis would prove lethal to mycobacterial cells. As with gal_f, rhamnose is neither synthesised nor utilised by mammalians thus it becomes an appealing target for the generation of new therapeutic agents. Similarly, as L- Rha biosynthesis is conserved across species (Ma *et al*, 1997) any new treatments would have a wide spectrum of activity.

The biosynthesis of dTDP - L - rhamnose from α - D – glucose–1–phosphate is catalysed by four enzymes Figure 6.1, (Tsukioka *et al*, 1997), termed RmlA, RmlB, RmlC and RmlD in a variety of both gram negative and positive bacteria including mycobacteria (Tsukioka *et al*, 1997; Ma *et al*, 1997; Brazeau *et al*, 1996). As a result, inhibition of any one of the four enzymes could prove effective in the inhibiting L – rha biosynthesis. The work described here involves the first two enzymes of the pathway, RmlA from *Streptococcus Suis* type II (*S. suis* II) and RmlB from *Salmonella enterica* (*S. enterica*).

Figure 6.1: Anabolism of dTDP-L-rhamnose from α-D-glucose-1-phosphate

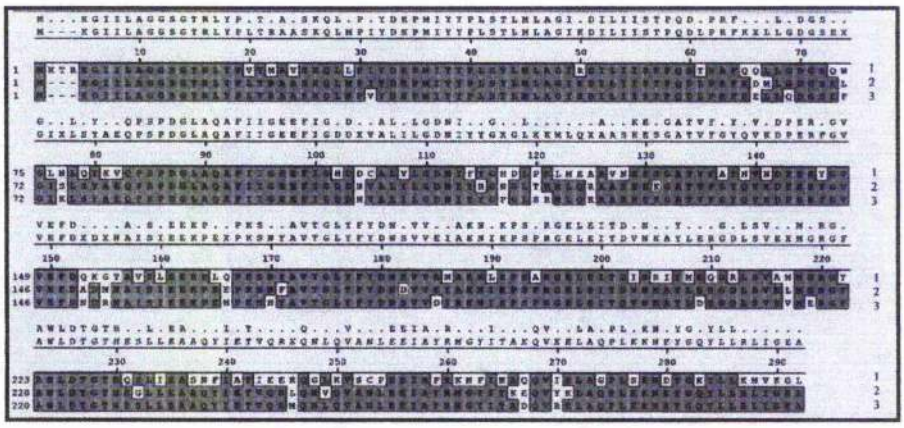


Glucose-1-Phosphate Thymidylyltransferase from Streptococcus suis Type II: RmlA

S. Suis II is a gram positive, haemolytic and facultative anaerobic pathogen. Presently 35 capsular serotypes of *S. Suis* exist with type II being the most commonly found in cases of disease and meningitis (Gottschalk *et al,* 1995). It is primarily a porcine pathogen where infection costs run to $300 million per annum in the United States (Staats *et al,* 1997). Human infection is rare but not unknown; slaughterhouse workers and butchers being the most at risk from infection (Tayaro *et al,* 1996). Ineffective vaccines and antimicrobials as a result of increased resistance by *S. suis II* has made the treatment of the disease difficult. This highlights the urgency with which the development of new drugs are sought. However, in designing inhibitors particular care will have to be exercised primarily to ensure that drugs are highly selective for RmlA and do not disrupt the similar reaction catalysed by glucose–1–phosphate uridylyltransferase where uridine 5'–phosphoryl is transferred from UDP–glucose to galactose–1–phosphate. This enzyme is essential in human metabolism.

RmlA catalyses the first step of the rhamnose pathway, the addition of dTDP and α-D–glucose–1–phosphate, Figure 6.1. The enzyme has been isolated from *S. suis II* and cloned in pET 21 for overexpression. Its molecular weight is 32 kDa and the pI is 4.7. At UV_{280} 1 AU corresponds to 1.23mg ml^{-1}. The primary sequence contains 10 methionines. As the enzyme is homologous between pathogens the solution of one three-dimensional structure by *de novo* methods will simplify the solution of subsequent structures, Figure 6.2.

Figure 6.2: Sequence Alignment of RmlA from (1)E. coli (2) S. suis II and (3) S. enterica

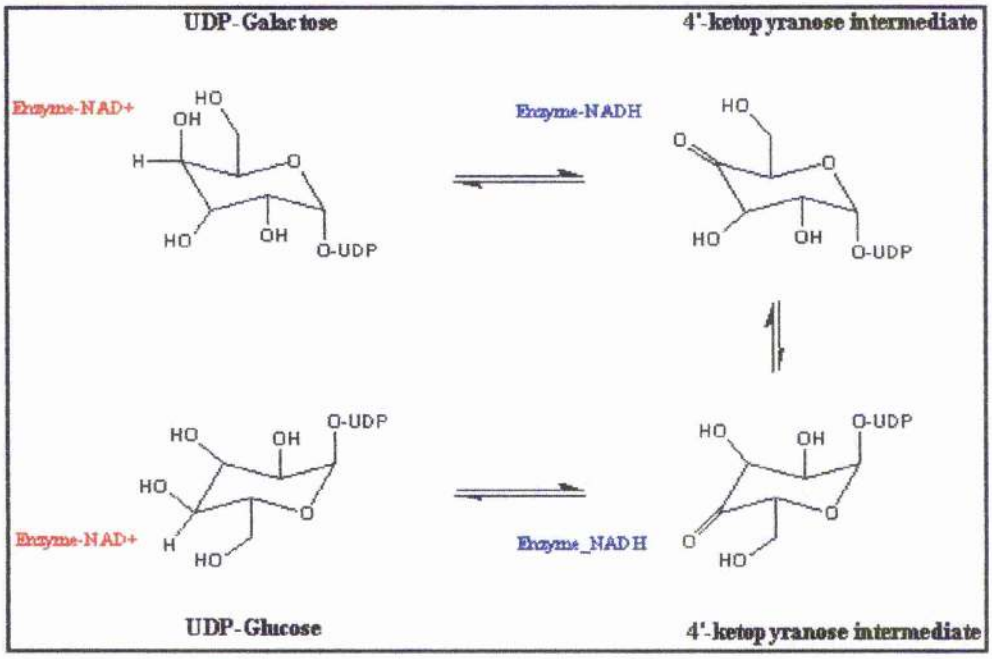


dTDP- D-glucose-4, 6-dehydratase from Salmonella enterica: RmlB

The genera Salmonella comes under the classification umbrella of enterobacteriaceae, implying a gram negative, aerobic and facultative anaerobic bacteria. It is the causative agent in many diseases including infections of the intestinal tract, the urinary tract septicemia and meningitis (Elliot *et al,* 1996).

Analysing the primary sequence of RmlB identifies 3 methionine residues, however unlike RmlA, RmlB will not require *de novo* phase acquisition as the structure of a 38% homologous protein UDP-galactose-4-epimerase (UDPGE) already exists. UDPGE catalyses inter conversion of UDP-galactose and UDP-glucose in normal galactose metabolism and the three dimensional structure has been elucidated to 1.8Å by Thoden & Holden (1998). The protein folds in two distinct motifs; an N-terminal or nucleotide binding domain consisting of a 7 stranded β sheet flanked by α helices. The smaller C-terminal domain consists of 3 β sheets and 2 α helices. The active site is wedged between the two domains. It is postulated the conversion proceeds through a 4'-ketopyranose intermediate where NAD^+ plays a crucial functional role in the mechanism, as shown in Figure 6.3.

Figure 6.3: Postulated mechanism for the interconversion of UDP-galactose and UDP-glucose (Thoden and Holden, 1998)



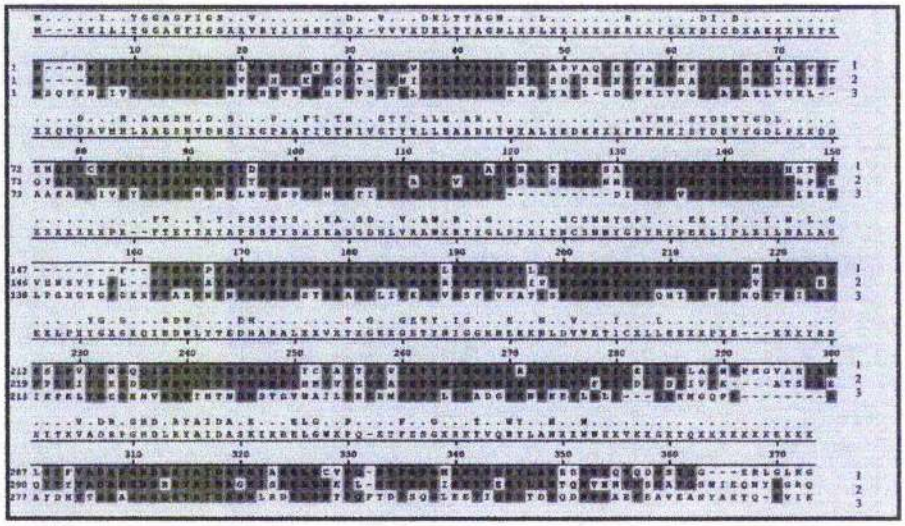
The key features of the mechanism include removal of the 4' hydroxyl by a catalytic base (Ser 124 and Tyr 149), the transfer of a hydride from C4 to NAD^+ and the rotation of the resulting 4'-ketopyranose in the active site which allows return of the hydride from NADH to the opposite face of the sugar. As the catalytic triad reported to be essential for UDPGE activity, Ser 124, Tyr 149 and Lys 153, is also present in RmlB, a similar mechanism of enzymatic action seems likely. The sequence alignment in Figure 6.4 illustrates the conserved amino acid regions from the epimerase discussed here, RmlB from *S. enterica* and RmlB from *Salmonella typhimurium*.

RmlB (TDP–D–glucose 4, 6-dehydratase) catalyses the conversion of dTDP–D–glucose to dTDP–6–deoxy-D-xylo–4–hexulose, Figure 6.1, and has been cloned from *S. enterica* ready for overexpression. The enzyme has a molecular weight of 42kDa and the point of no net

charge is 5.7. At UV_{280} 1 AU corresponds to 0.62mg ml^{-1}. Enzyme activity drops to below 10% when the protein is assayed in the absence of NAD^+. Thoden and Holden (unpublished, 1998) have also solved the structure of *E. coli* RmlB (PDB Accession Number 1BXK).

The overexpression, purification and crystallisation of native RmlB from *S. enterica* had previously been optimised by a colleague, Dr. Marie-France Giraud. Only work on the selenomethionine variant enzyme is discussed in this chapter. At the time of the work described in this chapter, the epimerase structure had not been reported, and consequently it was thought phases would have to have calculated *ab initio*. For this reason the experiments reported here were performed with a view to solving the three dimensional structure by MAD techniques.

Figure 6.4: Alignment report of (1) UDP-galactose-4-epimerase and RmlB (2) S. typhimurium and (3) S. suis II



6.3 Experimental

Overexpression of Recombinant RmlA form Streptococcus suis

All materials used were obtained from Sigma (Poole. UK) unless otherwise stated. Prior to all cultivation procedures, all containers, pipette tips and media were sterilised by autoclaving at 120°C for 30 minutes. Dr. Andrew Allan, University of Cambridge generously supplied the clone for RmlA.

In a typical experiment 5 milliliters of LB containing kanamycin (80μg ml^{-1}) was inoculated with a single colony of BL21 (DE3) cells containing the gene for RmlA and incubated overnight at 37°C. This culture was used to inoculate 500ml of similarly treated LB medium and the resultant solution shaken at 200rpm, 37°C. The optical density was monitored until the A_{600} reached 0.6 where upon IPTG was added to a final concentration of 1mM. The incubation was continued for a further 3 hours under the same conditions. The cells were then harvested by centrifugation (10000g, 4°C, 30 minutes) and frozen at –78°C until required.

Preparation of Cell Extracts.

Frozen cell paste was thawed in ice-cold lysis buffer (50mM TRIS pH 7.6, 1mM DTT, 1mM PMSF) at a concentration of 0.5g wet cell paste for each mL buffer. Lysozyme (200μg ml^{-1}), DNase (20μg ml^{-1}) and EDTA (2mM) were added and the solution left at room temperature for 30 minutes before the cell contents were released by sonication (140kW, 20kHz) at 4°C for 6 x 30 seconds, resting on ice in between. Cell debris was pelleted by centrifugation (20000g, 4°C, 30 minutes). The cell extract was transferred to a clean tube and the contents assayed by the Bradford assay (Bradford, 1976) and SDS-PAGE.

Ammonium Sulphate Fractionation and Anion Exchange Chromatography

Cell extract was brought to 10% $(NH_4)_2SO_4$ saturation by the slow addition of 5.6g $(NH_4)_2SO_4$ per $100ml^{-1}$ cell extract with gentle stirring at 4°C. Precipitated contaminants were pelleted by centrifugation and the supernatant dialysed 3 times against a 1000ml solution of 50mM TRIS pH7.5, 1mM DTT to remove ammonium sulphate before application to an anion exchange chromatography column. Dialysed protein was 0.02μm filtered and applied to a 7.84ml anion exchange column (Perseptive Biosystems, HQ) equilibrated in buffer A (50mM TRIS pH 7.6, 1mM DTT) in 100mg aliquots at $5ml\ min^{-1}$. The optimum pH for separation was determined experimentally through application of a pH multi-method. Small amounts of protein were applied to the column and eluted at pH levels differing in 0.5 units over the pH range 6-9. Eluent was collected over 15 column volumes (CV's) by applying a 0-500mM NaCl gradient. Protein elution was monitored on-line at 280nm. The required fraction was identified after analysis by SDS – PAGE and appropriate peaks from subsequent runs were pooled and concentrated under nitrogen pressure in an Amicon ultrafiltration cell.

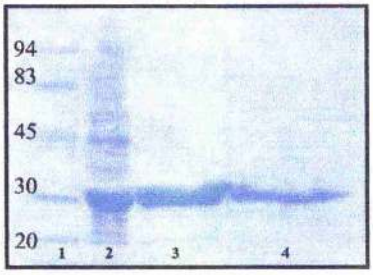
Hydrophobic Interaction Chromatography.

Pooled fractions from the anion exchange column were brought to 35% $(NH_4)_2SO_4$ saturation by dialysis overnight against a solution of 20mM TRIS, 35% $(NH_4)_2\ SO_4$ and 1mM DTT. The protein solution was 0.02μm filtered again before application at $5ml\ min^{-1}$ to a 7.84ml hydrophobic interaction chromatography column (Perseptive Biosystems, HP) equilibrated in buffer B (50mM TRIS pH 7.3, 35% $(NH_4)_2SO_4$ 1mM DTT). Protein was

eluted in a decreasing $(NH_4)_2SO_4$ gradient (100-0%) in buffer B over 15 CV's. As before the correct fraction was identified by SDS – PAGE, Figure 6.5.

Protein concentration was assayed by the Bradford method (Bradford, 1976) and UV absorption at 280nm.

Figure 6.5: Coomassie blue stained SDS-PAGE displaying protein purity after subsequent purification steps.
Lane 1: Molecular markers (kDa)
Lane 2: Crude Rml A.
Lane 3: Post HQ Rml A.
Lane 4: Post HP Rml A.



The protein ran as a single band on SDS-PAGE and was judged to be satisfactorily pure for crystallisation.

Protein was also overexpressed in terrific broth as opposed to LB. However, as the yield of soluble protein obtained was essentially the same (11.2 mgs from terrific broth and 10.4 mgs from LB) per litre of culture over expression was routinely performed in LB for cost effectiveness.

Dynamic Light Scattering

RmlA was concentrated to 5mg ml^{-1} under nitrogen pressure (Amicon) and dialysed 3 times against 20mM TRIS pH 7.3, 2mM DTT in preparation for analysis by dynamic light scattering. The results of which are shown in Table 6.1.

Table 6.1: Dynamic light scattering results for Rml A. The average polydispersity value as a function of the hydrodynamic radius is 30.5%.

#	Ampl.	Differ. Coeff.	Radius (nm)	Polyd. (nm)	Est.MW kDa	Temp. ^{0}C	Count Rate	Base Line	Poly D%
1	0.807	533	4.2	1.450	96	22.3	109	1.023	34.5
2	0.827	547	4.1	1.197	91	21.8	105	1.003	29.1
3	0.815	522	4.2	1.242	99	22.1	109	1.023	29.5
4	0.806	514	4.3	1.268	104	22.3	110	1.029	29.4
5	0.823	512	4.4	1.517	100	22.1	109	1.006	35.2
6	0.835	520	4.4	1.551	106	22.3	109	1.015	35.2
7	0.837	504	4.2	1.744	109	22.2	110	1.021	41.5
8	0.844	533	4.3	1.476	96	22.4	106	1.005	35.1
9	0.846	522	4.2	1.513	100	22.2	107	1.004	36.0
10	0.842	540	4.1	1.205	92	22.2	105	1.003	29.3

Although the average polydispersity as a function of the hydrodynamic radius is slightly higher than one would like for crystallisation purposes, protein was routinely produced by this protocol and used in crystallisation trials.

Crystallisation of Rml A

Unlike the crystallisation of UDP-galactopyranose mutase, the sitting drop method of crystallisation was employed for initial crystallisation trials. Rather than 'hanging' in a sealed environment in the presence of the crystallising agent the protein 'sits' in an elevated well above the crystallising agent. The reason for the switch in approach was for purposes of convenience only. The plates used for this set up are available from Douglas Instruments and enable a more cost efficient approach to crystallisation.

After the final purification step protein was concentrated to 10 mg ml^{-1} and dialysed against three changes of 20mM TRIS pH 7.3, 2mM DTT. Crystallisation trials were performed with 2µl of protein solution and 2µl of precipitant and incubated at 20.5°C. As before the

crystallisation screens available from Hampton Research were used to screen for primary crystallisation conditions. All conditions tried are listed in Table 6.2, along with the results observed.

Table 6.2: Initial screens performed in crystallisation of Rml A.

Precipitant	**pH**	**Protein Concentration mg ml^{-1}**	**Mixing**	**Crystals**
PEG 6K	4, 5, 6, 7, 8, & 9	2, 4 & 8	Yes	No
PEG 6K / LiCl				Yes
$(NH_4)_2SO_4$				Yes
MPD				No
NaCl				No
Hampton CS I				No
Hampton CS II				No

After 4 days small crystals (0.05 x 0.05 x 0.01mm) had grown from solutions equilibrated against 30 % PEG 6K / LiCl at pH units ranging from 6 – 9. To try and increase the size of the crystals the parameters around this condition were optimised. Thereafter the optimum size of crystal was grown from a solution of 4mg ml^{-1} protein equilibrated with 23% PEG / 1.0M LiCl, 0.1M HEPES pH 6.5.

Encouraging results were also obtained from the initial screens with $(NH_4)_2SO_4$. After 8 – 9 days crystal nucleation sites were evident. As with the PEG / LiCl initial conditions, optimisation followed and crystallisation parameters were refined allowing hexagonal crystals to grow from a protein solution equilibrated against 2.8M $(NH_4)_2SO_4$, 0.1M MES pH 6.2 after 9 days, Figure 6.6. However, crystal size was so small (0.05 x 0.05 x ~0.01mm) that continued optimisation of the PEG / LiCl conditions took priority.

Additive screens from Hampton Research were tried at this stage to see if they complimented crystal growth. Of the 72 different additives tested only $Mn(Cl_4)_2$ at 0.01M appeared to enhance crystal growth. This was subsequently included in crystallisation experiments. Incubation of the protein with substrate (α-glucose-1-phosphate) at 4°C and 37°C prior to crystallisation did not promote the growth of larger crystals. The final size of the crystals was now 0.3 x 0.1 x 0.1mm and growth was complete after 4 days, Figure 6.7.

Figure 6.6: Crystal of RmlA grown from 2.8M $(NH_4)_2SO_4$, 0.1M MES pH 6.2

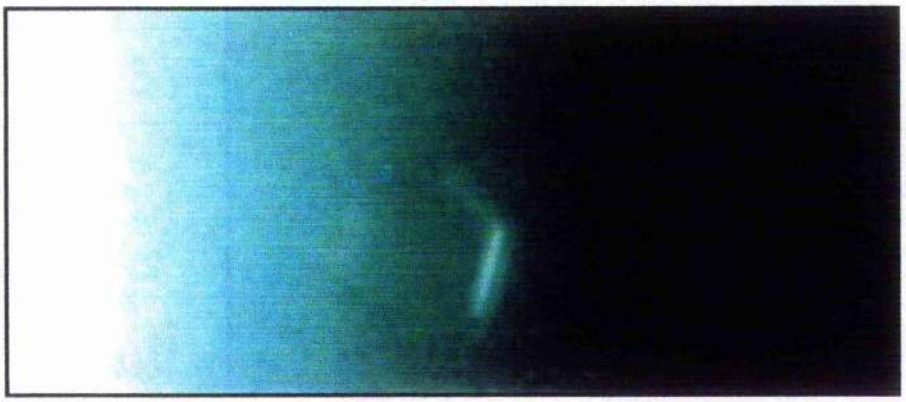


Figure 6.7: Crystals of RmlA grown from 23% PEG / 1.0M Li Cl, 0.01M Mn $(Cl_4)_2$, 0.1M HEPES pH6.5

Data Collection

A crystal was cryoprotected in a solution of mother liquor containing 15% glycerol and mounted in the cryostream (110K). The crystal to detector distance was set to 200mm and an exposure time of 50 minutes per 1° oscillation employed. X–rays were generated using a FR–951 rotating anode (Enraf Nonius) with copper target and focussed with the MacScience mirror system. As data was only observable to 8Å indexing were impossible. However, a single image was recorded on PX 9.6 at the synchrotron source Daresbury, which showed diffraction to 3.5Å.

Preparation of Selenomethionine Variant RmlB

DNA Extraction and Transformation into the Methionine Auxotroph B834 (DE3)

The method followed is exactly as outlined in Chapter 2, for the analogous experiment with UDP-galactopyranose mutase.

Overexpression and Purification of B834 RmlB in Defined Medium.

Previous experiments with mutase to quantify the amount of soluble protein obtained using different expression media for the overexpression of selenomethionine variant protein showed no particular medium to excel. For this reason the same media (defined) used in over-expression of mutase was used here.

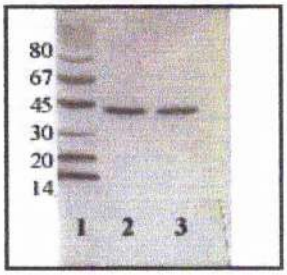
Single colonies were selected from an agar/ampicillin plate and used to inoculate 2 x 5ml of defined medium (Table 2.3) supplemented with 1mg L^{-1} thiamine. This starter culture was incubated at 37°C overnight with agitation and used to inoculate 2 x 500ml of defined

medium. Incubation continued at 37°C 200rpm before addition of 1mM IPTG at an OD_{600} of 1.0 whereupon incubation continued for a further 16 hours before the cells were harvested by centrifugation (10000*g*, 4°C, 15 minutes).

Purification of Selenomethionine Variant RmlB

Purification of the selenomethionine variant RmlB was as the native enzyme (Giraud personal communication). Experience in the purification of the selenomethionine variant mutase suggested a concentration of 2mM DTT would be sufficient in preventing oxidation. For this reason DTT concentrations were kept at this level to avoid any potential disruption of the purification strategy. Sample homogeneity was judged by SDS-PAGE, Figure 6.8.

Figure 6.8: Silver stained SDS-PAGE of Se-Met RmlB after hydrophobic interaction chromatography. Lane1 molecular markers (kDa), lanes 2&3 purified RmlB.



Mass Spectrometric Analysis of RmlB

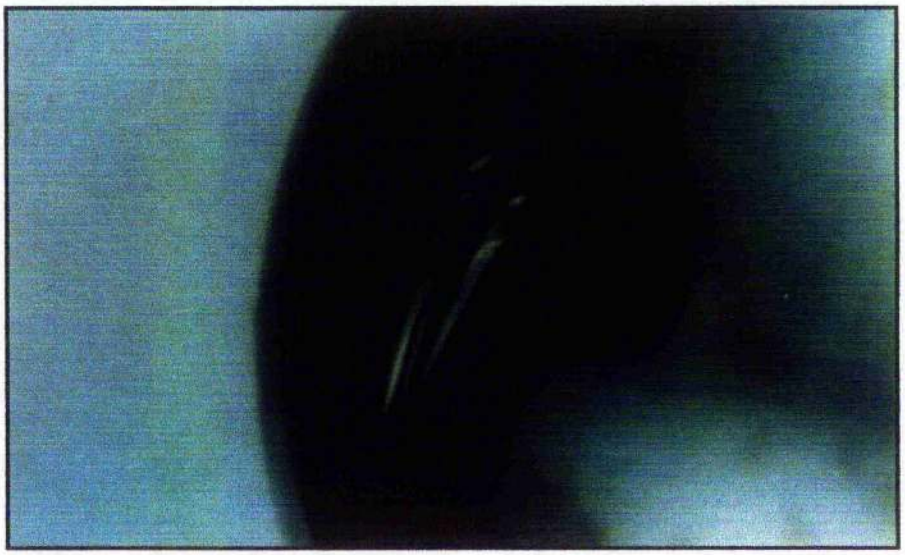
Selenium incorporation was confirmed by mass spectrometry to be 100%. Analysis of the selenomethionine protein revealed a peak at 40834 a.m.u, 154 a.m.u. higher than the peak displayed by native protein, which is correct for the substitution of three selenium's in methionine.

Crystallisation of Selenomethionine Variant RmlB.

As the crystallisation parameters for the native protein had been previously determined and optimised (0.1M MES pH 6.3, 1.5M $LiSO_4$, 2mM DTT) screens around this condition were

set up with selenomethionine RmlB in sitting drop trays. As with native enzyme, the protein (3mg ml^{-1}) was incubated prior to crystallisation overnight with NAD^+. Crystal growth was complete after 14 days. The maximum crystal size was smaller than that for the native enzyme at 0.4 x 0.15 x 0.15mm compared to 0.6 x 0.2 x 0.2 mm, Figure 6.9.

Figure 6.9: Selenomethionine RmlB crystallised from a solution of 3mg ml^{-1} protein equilibrated against 0.1M MES pH 6.3, 1.5M $LiSO_4$, 2mM DTT. Crystal dimensions 0.4 x 0.15 x 0.15mm.



In-house Selenomethionine Variant RmlB Data Collection

A crystal of optimum size was cryoprotected in 15% glycerol and mounted in the cryostream at 110K. All diffraction data were collected using the Enraf-Nonius/MacScience DIP2000 dual image plate. X-rays were generated (λ=1.54Å) using an Enraf-Nonius FR591 rotating anode generator and focussed using the MacScience mirror system. An oscillation range of 1° was set, exposure time 25 minutes and the crystal to detector distance set at 150mm. The crystal diffracted weakly to 3.1Å. As a result the exposure time was doubled and the image plate moved back to 200mm, 3.0Å at edge.

90 degree's of data were collected and processed using DENZO and SCALEPACK (Otwinowski & Minor, 1996), in space group $C222_1$ with unit cell dimensions a = 146.62Å, b = 168.92Å, c = 87.72Å, $\alpha = \beta = \gamma = 90^0$. The data are 85% complete from 26Å to 3.1Å, R_{merge} 11.9% and redundancy 2.6. Unfortunately the crystal iced during data acquisition preventing the collection of a more complete data set. Data collection and reduction statistics are shown in Table 6.3. The mosaic spread was fixed to 0.5° for scaling, as it proved unstable in refinement.

Table 6.3: Quality of the selenomethionine variant protein data set of RmlB collected in-house.

Resolution (Å)	No. Reflections	% Complete	R-Merge %	Redundancy
40.00 – 6.67	1753	81.2	4.8	2.7
6.67 – 5.30	1784	86.0	9.7	2.7
5.30 – 4.63	1786	87.3	8.6	2.7
4.63 – 4.21	1791	88.2	9.2	2.7
4.21 – 3.91	1746	85.5	12.5	2.6
3.91 – 3.68	1717	85.7	15.3	2.6
3.68 – 3.49	1760	87.6	18.7	2.6
3.49 – 3.34	1752	86.6	22.7	2.6
3.34 – 3.21	1749	88.1	27.9	2.5
3.21 – 3.10	1433	71.6	35.0	2.5
40.00 – 3.10	**17271**	**84.8**	**11.9**	**2.6**

6.4 Results and Discussion

RmlA

RmlA from *Streptococcus suis* has successfully been overexpressed and purified in a form that is suitable for protein crystallisation. The protein migrates as a single band around 30 kDa on denaturing gels. However, analysis by light scattering suggests a certain amount of heterogeneity may be present. Regardless, crystals have been reproducibly grown from solutions of PEG / LiCl and $(NH_4)_2SO_4$ by the sitting drop vapour diffusion method. The

maximum crystal size currently obtainable is 0.3 x 0.1 x 0.1mm. Characterisation of the crystals by X-ray analysis shows diffraction to around 8Å in-house and 3.5Å on PX 9.6, Daresbury.

To obtain higher resolution data, crystals of better quality are required. Further optimisation of the current crystallisation conditions may yield better crystals and this should be pursued particularly with the ammonium sulphate conditions, which have not had much optimisation as yet. Alternatively, as analysis by dynamic light scattering shows some polydispersity, the purification protocol may require refinement. Before alteration of the purification protocol the protein should first be characterised by mass spectrometry. If any heterogeneity is apparent then optimisation of the purification protocol will then become the priority.

RmlB

A selenomethionine variant of the protein has also been successfully overexpressed, purified and crystallised in a form suitable for X-ray data collection. Crystal growth characteristics are identical to the native protein, though the final crystal dimensions are approximately 25% smaller. 100% selenium incorporation has been confirmed by mass spectrometry making the crystals suitable for MAD analysis. A 85% data set, R_{merge} = 11.9%, has been collected in-house using one of these crystals. As the structure of a 40% homologous structure has since been deposited in the PDB the need for a selenomethionine variant of the protein has been removed.

6.5 Conclusions and Future Work

RmlA

Parameters allowing the reproducible crystallisation of RmlA have been determined and improvement of crystal quality is the initial priority. Should crystals of higher quality be forthcoming then data collection should commence. As there is no protein structure currently exhibiting significant sequence similarity in the protein data bank (http://www.rcsb.org), phases will have to be determined experimentally. Transformation of the gene encoding for RmlA into B834 cells will allow overexpression of the selenomethionine variant of the protein where a high level of selenium incorporation would be expected. Purification and crystallisation should be highly similar to that of the native protein and yield crystals where selenium has replaced sulphur in the protein. Phases will then be determined from MAD methods.

RmlB

This structure is now likely to be solved by molecular replacement and others in the laboratory are actively pursuing this.

Appendices

Appendix I

Hampton Crystal Screen I

Tube	Salt	Tube	Buffer	Tube	Precipitate
1.	0.02M $CaCl_2$	1.	Na Acetate	1.	30% v/v MPD
2.	None	2.	None	2.	0.4M K, Na tartrate
3.	None	3.	None	3.	0.4M NH_4 Phosphate
4.	None	4.	0.1M TRIS pH8.5	4.	2.0M $(NH)_4 SO_4$
5.	0.2M Na Citrate	5.	0.1M Na HEPES pH7.5	5.	30%v/v MPD
6.	0.2M $MgCl_2$	6.	0.1M TRIS pH 8.5	6.	30%w/v PEG 4000
7.	None	7.	0.1M Na Cacodylate pH6.5	7.	1.4M Na Acetate
8.	0.2M Na Citrate	8.	0.1M Na Cacodylate pH6.5	8.	30%v/v 2-propanol
9.	0.2M NH_4 acetate	9.	0.1M Na Acetate pH 5.6	9.	30%w/v PEG4000
10.	0.2M NH_4 acetate	10.	0.1M Na Acetate pH 4.6	10.	30%w/v PEG 4000
11.	None	11.	0.1M Na Citrate pH 5.6	11.	1.0M NH_4 Phosphate
12.	0.2M $MgCl_2$	12.	0.1M Na HEPES pH 7.5	12.	30%v/v 2-propanol
13.	0.2M Na Citrate	13.	0.1M TRIS pH 8.5	13	30%v/v PEG 400
14.	0.2M $CaCl_2$	14.	0.1M Na HEPES pH 7.5	14.	28%v/v PEG 400
15.	0.2M $(NH)_4 SO_4$	15.	0.1M Na Cacodylate pH6.5	15.	30%w/v PEG8000
16.	None	16.	0.1M Na HEPES pH 7.5	16.	1.5M $Li SO_4$
17.	0.2M $Li SO_4$	17.	0.1M TRIS pH 8.5	17.	30% PEG 4000
18.	0.2M Mg Acetate	18.	0.1M Na Cacodylate pH6.5	18.	20% PEG 8000
19.	0.2M NH_4 Acetate	19	0.1M TRIS pH 8.5	19.	30%v/v 2-propanol
20.	0.2M $(NH)_4 SO_4$	20.	0.1M Na Acetate pH 4.6	20.	25% w/v PEG 4000
21.	0.2M Mg Acetate	21.	0.1M Na Cacodylate pH6.5	21.	30%v/v MPD
22.	0.2M Na Acetate	22.	0.1M TRIS pH 8.5	22.	30%w/v PEG 4000
23.	0.2M $MgCl_2$	23.	0.1M Na HEPES pH 7.5	23.	30%v/v PEG 400
24.	0.2M $CaCl_2$	24.	0.1M Na Acetate pH4.6	24.	20%v/v 2-propanol
25.	None	25.	0.1M Imidazole pH 6.5	25.	1.0M Na Acetate
26.	0.2M NH_4 Acetate	26.	0.1M Na Citrate pH 5.6	26.	30%v/v MPD
27.	0.2M Na Citrate	27.	0.1M Na HEPES pH 7.5	27.	20%v/v 2-propanol
28.	0.2M Na Acetate	28.	0.1M Na Cacodylate pH6.5	28.	30%w/v PEG 8000
29.	None	29.	0.1M Na HEPES pH 7.5	29.	0.8M K, Na Tartrate
30.	0.2M $NH_4 SO_4$	30.	None	30.	30%w/v PEG 8000
31.	0.2M $NH_4 SO_4$	31.	None	31.	30%w/v PEG 4000
32.	None	32.	None	32.	2.0M $(NH)_4 SO_4$
33.	None	33.	None	33.	4.0M Na Formate
34.	None	34.	0.1M Na Acetate pH 4.6	34.	2.0M Na Formate
35.	None	35.	0.1M Na HEPES pH 7.5	35.	1.6M Na, K Phosphate
36.	None	36.	0.1M TRIS pH 8.5	36.	8%w/v PEG8000
37.	None	37.	0.1M Na Acetate pH 4.6	37.	8%w/v PEG4000
38.	None	38.	0.1M Na HEPES pH 7.5	38.	1.4M Na Citrate
39.	None	39.	0.1M Na HEPES pH 7.5	39.	2%v/v PEG 400 & 2.0M NH_4 SO4
40.	None	40.	0.1M Na Citrate pH5.6	40.	20%v/v 2-propanol & 20% PEG 400
41.	None	41.	0.1M Na HEPES pH 7.5	41.	10% v/v 2-propanol & 20% PEG 4000
42.	0.05M K Phosphate	42.	None	42.	20w/v PEG 8000
43.	None	43.	None	43.	30%v/v PEG 1500

44. None	44.	None	44.	0.2M Mg Acetate
45. 0.2M Zn Acetate	45.	0.1M Na Cacodylate pH6.5	45.	18%w/v PEG 8000
46. 0.2M Ca Acetate	46.	0.1M Na Cacodylate pH6.5	46.	18%w/v PEG 8000
47. None	47.	0.1M Na Acetate pH 4.6	47.	2.0M $(NH)_4$ SO_4
48. None	48.	0.1M TRIS pH 8.5	48.	2.0M $(NH)_4$ SO_4
49. 1.0M Li SO_4	49.	None	49.	2%w/v PEG 8000
50. 0.5M Li SO_4	50.	None	50.	15%w/v PEG 8000

Hampton Crystal Screen II

Tube	Salt	Tube	Buffer	Tube	Precipitate
1.0	2M NaCl	1.0.	None	1.0	10%w/v PEG 6000
2.0	0.1M acetyl$(CH_3)_3$NHBr	2.0	None	2.0	0.5M NaCl 0.01M $MgCl_2.(OH)_6$
3.0	None	3.0	None	3.0	25% v/v Ethylene glycol
4.0	None	4.0	None	4.0	35% v/v Dioxane
5.0	2.0M $(NH)_4$ SO_4	5.0	None	5.0	5% v/v isopropanol
6.0	None	6.0	None	6.0	1.0M Imidazole pH 6.5
7.0	None	7.0	None	7.0	10% w/v PEG 1000 10% w/v PEG 8000
8.0	1.5M NaCl	8.0	None	8.0	10% v/v Ethanol
9.0	None	9.0	0.1M Na Acetate pH 4.6	9.0	2.0M Na Cl
10.0	0.2M NaCl	10.0	0.1M Na Acetate pH 4.6	10.0	30% v/v MPD
11.0	0.01M $CoCl.(OH)_6$	11.0	0.1M Na Acetate pH 4.6	11.0	1.0M 1,6 Hexanediol
12.0	0.1M $CdCl_2$	12.0	0.1M Na Acetate pH 4.6	12.0	30% w/v PEG 400
13.0	0.2M $(NH)_4$ SO_4	13.0	0.1M Na Acetate pH 4.6	13.0	30% w/v PEG monomethylether 2000
14.0	0.2M K/Na Tartrate	14.0	0.1M Na Citrate pH 5.6	14.0	2.0M $(NH)_4$ SO_4
15.0	0.5M $(NH)_4$ SO_4	15.0	0.1M Na Citrate pH 5.6	15.0	1.0M Li_2 SO_4
16.0	0.5M NaCl	16.0	0.1M Na Citrate pH 5.6	16.0	4% w/v Polyethyleneimine
17.0	None	17.0	0.1M Na Citrate pH 5.6	17.0	35% v/v tert-butanol
18.0	0.01M $FeCl_2.(OH)_6$	18.0	0.1M Na Citrate pH 5.6	18.0	10% v/v Jeffamine M-600
19.0	None	19.0	0.1M Na Citrate pH 5.6	19.0	2.5M 1,6 Hexanediol
20.0	None	20.0	0.1M MES pH 6.5	20.0	1.6M $MgSO_4.(OH)_7$
21.0	0.1M $NaPO_4$ 0.1M KPO_4	21.0	0.1M MES pH 6.5	21.0	2.0M NaCl
22.0	None	22.0	0.1M MES pH6.5	22.0	12% w/v PEG 20,000
23.0	1.6M $(NH)_4$ SO_4	23.0	0.1M MES pH6.5	23.0	10% v/v Dioxane
24.0	0.05M CsCl	24.0	0.1M MES pH 6.5	24.0	30% Jeffamine M-600
25.0	0.01M $CoCl.(OH)_6$	25.0	0.1M MES pH 6.5	25.0	1.8M $(NH4)_2SO_4$
26.0	0.2M $(NH4)_2SO_4$	26.0	0.1M MES pH 6.5	26.0	30% w/v PEG monomethylether 5000
27.0	0.01M $ZnSO_4.(OH)_7$	27.0	0.1M MES pH 6.5	27.0	25% v/v PEG monomethylether 550
28.0	None	28.0	None	28.0	1.6M Na Citrate pH6.5
29.0	0.5M $(NH_4)_2SO_4$	29.0	0.1M HEPES pH 7.5	29.0	30% v/v MPD
30.0	None	30.0	0.1M HEPES pH 7.5	30.0	10% w/v PEG 6000 5% v/v MPD

31.0	None	**31.0**	0.1M HEPES pH 7.5	**31.0**	20% v/v Jeffamine M-600
32.0	0.1M NaCl	**32.0**	0.1M HEPES pH 7.5	**32.0**	1.6M $(NH4)_2SO_4$
33.0	None	**33.0**	0.1M HEPES pH 7.5	**33.0**	2.0M (NH4) Formate
34.0	0.05MCd $SO_4.(OH)_8$	**34.0**	0.1M HEPES pH 7.5	**34.0**	1.0M Na Acetate
35.0	None	**35.0**	0.1M HEPES pH 7.5	**35.0**	70% v/v MPD
36.0	None	**36.0**	0.1M HEPES pH 7.5	**36.0**	4.3M NaCl
37.0	None	**37.0**	0.1M HEPES pH 7.5	**37.0**	10% w/v PEG 8000 8% v/v Ethylene glycol
38.0	None	**38.0**	0.1M HEPES pH 7.5	**38.0**	20% w/v PEG 10000
39.0	0.2M $MgCl_2.(OH)_7$	**39.0**	0.1M TRIS pH 8.5	**39.0**	3.4 M 1.6 Hexanediol
40.0	0.1M $CaCl_2.(OH)_2$	**40.0**	0.1M TRIS pH 8.5	**40.0**	25% v/v tert-butanol
41.0	0.01M NiCl. $(OH)_6$	**41.0**	0.1M TRIS pH 8.5	**41.0**	1.0M Li2SO4
42.0	1.5M $(NH_4)_2SO_4$	**42.0**	0.1M TRIS pH 8.5	**42.0**	12% v/v Glycerol
43.0	0.2M $(NH_4)PO_3$	**43.0**	0.1M TRIS pH 8.5	**43.0**	50% v/v MPD
44.0	None	**44.0**	0.1M TRIS pH 8.5	**44.0**	20% v/v Ethanol
45.0	0.01M NiCl. $(OH)_6$	**45.0**	0.1M TRIS pH 8.5	**45.0**	20% w/v PEG monomethylether 2000
46.0	0.1M NaCl	**46.0**	0.1M Bicine pH 9.0	**46.0**	30% w/v PEG monomethyether 550
47.0	None	**47.0**	0.1M Bicine pH9.0	**47.0**	2.0M $MgCl_2.(OH)_6$
48.0	2% v/v dioxane	**48.0**	0.1M Bicine pH 9.0	**48.0**	10% w/v PEG20000

JMac Crystal Screen

1.0	0.05M $CaCl_2$	**1.0**	0.1M Acetate pH4.6	**1.0**	15% MPD
2.0	0.1M Li_2SO_4	**2.0**	0.1M Acetate pH4.6	**2.0**	10% PEG 8K
3.0	0.2M Mg Cl_2	**3.0**	0.1M Acetate pH4.6	**3.0**	30% PEG 8K
4.0	None	**4.0**	0.1M Acetate pH4.6	**4.0**	1M Li_2SO_4
5.0	0.5M $(NH_4)_2SO_4$	**5.0**	0.1M Acetate pH4.6	**5.0**	10% PEG 8K
6.0	0.5M MgCl2	**6.0**	0.1M Acetate pH4.6	**6.0**	2M $(NH_4)_2SO_4$
7.0	0.2M NaK Phosphate	**7.0**	0.1M Acetate pH4.6	**7.0**	3M $(NH_4)_2SO_4$
8.0	None	**8.0**	0.1M Acetate pH4.6	**8.0**	1M Na tartrate
9.0	0.1M Na tartrate	**9.0**	0.1M Acetate pH 4.6	**9.0**	20% PEG 8K
10.0	None	**10.0**	0.1M Acetate pH 4.6	**10.0**	2M NaCl
11.0	0.1M Na tartrate	**11.0**	0.1M Acetate pH 4.6	**11.0**	10% MPD
12.0	0.2M Li_2SO_4	**12.0**	0.1M Acetate pH 4.6	**12.0**	20% PEG 8K
13.0	0.2M NaCl	**13.0**	0.1M Acetate pH 4.6	**13.0**	30% MPD
14.0	None	**14.0**	0.1M Acetate pH 4.6	**14.0**	23M Na / K phosphate
15.0	0.2M Li_2SO_4	**15.0**	0.1M Na Citrate pH 5.6	**15.0**	20% PEG 8K
16.0	None	**16.0**	0.1M Na Citrate pH 5.6	**16.0**	30% PEG 8K
17.0	0.2M Na tartrate	**17.0**	0.1M Na Citrate pH 5.6	**17.0**	2M $(NH_4)_2SO_4$
18.0	None	**18.0**	0.1M Na Citrate pH 5.6	**18.0**	10% PEG 8K
19.0	0.5M $(NH_4)_2SO_4$	**19.0**	0.1M Na Citrate pH 5.6	**19.0**	10% PEG 8K
20.0	0.2M Na Cl	**20.0**	0.1M Na Citrate pH 5.6	**20.0**	3M Na / K phosphate
21.0	0.2M $MgCl_2$	**21.0**	0.1M Na Citrate pH 5.6	**21.0**	20% PEG 8K
22.0	0.2M Mg Cl2	**22.0**	0.1M Na Citrate pH 5.6	**22.0**	1M Na tartrate

23.0	0.2M Na/K phosphate	**23.0**	0.1M Na Citrate pH 5.6	**23.0**	2M $(NH_4)_2$ SO_4
24.0	None	**24.0**	0.1M Na Citrate pH 5.6	**24.0**	2M Na Cl
25.0	0.5M $(NH_4)_2$ SO_4	**25.0**	0.1M Na Citrate pH 5.6	**25.0**	1M Li_2 SO_4
26.0	0.1M Mg Cl_2	**26.0**	0.1M Na Citrate pH 5.6	**26.0**	3M Na/K phosphate
27.0	0.2M Li_2SO_4	**27.0**	0.1M Na Citrate pH 5.6	**27.0**	3.0M $(NH_4)_2$ SO_4
28.0	0.2M Na/K phosphate	**28.0**	0.1M Na Citrate pH 5.6	**28.0**	30% MPD
29.0	None	**29.0**	0.1M Na Citrate pH 5.6	**29.0**	30% v/v MPD
30.0	None	**30.0**	0.1M MOPS pH6.5	**30.0**	20% PEG 8K
31.0	None	**31.0**	0.1M MOPS pH6.5	**31.0**	1M Na tartrate
32.0	0.2M Na tartrate	**32.0**	0.1M MOPS pH6.5	**32.0**	2M Na Cl
33.0	0.2M Li_2 SO_4	**33.0**	0.1M MOPS pH6.5	**33.0**	3M Na/K phosphate
34.0	0.2M NaCl	**34.0**	0.1M MOPS pH6.5	**34.0**	20% PEG 8K
35.0	None	**35.0**	0.1M MOPS pH6.5	**35.0**	2M $(NH_4)_2$ SO_4
36.0	None	**36.0**	0.1M MOPS pH6.5	**36.0**	3.0M $(NH_4)_2$ SO_4
37.0	0.2M Na/K phosphate	**37.0**	0.1M MOPS pH6.5	**37.0**	15% PEG 8K
38.0	0.2M Na tartrate	**38.0**	0.1M MOPS pH6.5	**38.0**	10% PEG 8K
39.0	0.5M $(NH_4)_2$ SO_4	**39.0**	0.1M MOPS pH6.5	**39.0**	20% PEG 8K
40.0	None	**40.0**	0.1M MOPS pH6.5	**40.0**	30% MPD
41.0	None	**41.0**	0.1M MOPS pH6.5	**41.0**	0.2M $CaCl_2$
42.0	0.2M Na/K phosphate	**42.0**	0.1M MOPS pH6.5	**42.0**	2M NaCl
43.0	0.2M $MgCl_2$	**43.0**	0.1M MOPS pH6.5	**43.0**	30% MPD
44.0	None	**44.0**	0.1M MOPS pH6.5	**44.0**	1M Li_2 SO_4
45.0	None	**45.0**	0.1M HEPES pH7.5	**45.0**	1M Li_2 SO_4
46.0	0.1M $CaCl_2$	**46.0**	0.1M HEPES pH7.5	**46.0**	10% PEG 8K
47.0	0.2M NaCl	**47.0**	0.1M HEPES pH7.5	**47.0**	3M $(NH_4)_2$ SO_4
48.0	None	**48.0**	0.1M HEPES pH7.5	**48.0**	20% PEG 8K
49.0	0.2M Na tartrate	**49.0**	0.1M HEPES pH7.5	**49.0**	3M $(NH_4)_2$ SO_4
50.0	0.2M Li_2SO_4	**50.0**	0.1M HEPES pH7.5	**50.0**	2M $(NH_4)_2$ SO_4
51.0	0.5M $(NH_4)_2$ SO_4	**51.0**	0.1M HEPES pH7.5	**51.0**	30% MPD
52.0	0.2M Na tartrate	**52.0**	0.1M HEPES pH7.5	**52.0**	3M Na/K phosphate
53.0	0.2MNa/K phosphate	**53.0**	0.1M HEPES pH7.5	**53.0**	1M Na tartrate
54.0	None	**54.0**	0.1M HEPES pH7.5	**54.0**	15% PEG 8K
55.0	None	**55.0**	0.1M HEPES pH7.5	**55.0**	2M NaCl
56.0	0.2M $MgCl_2$	**56.0**	0.1M HEPES pH7.5	**56.0**	19% PEG 8K
57.0	0.2M Li_2SO_4	**57.0**	0.1M HEPES pH7.5	**57.0**	20% PEG 8K
58.0	0.2M Na tartrate	**58.0**	0.1M HEPES pH7.5	**58.0**	1M Li_2 SO_4
59.0	0.2M Li_2SO_4	**59.0**	0.1M TRIS pH8.5	**59.0**	2M NaCl
60.0	0.2M Na/Kphosphate	**60.0**	0.1M TRIS pH8.5	**60.0**	20% PEG 8K
61.0	0.2M NaCl	**61.0**	0.1M TRIS pH8.5	**61.0**	20% PEG 8K
62.0	0.2M NaCl	**62.0**	0.1M TRIS pH8.5	**62.0**	2M $(NH_4)_2$ SO_4
63.0	0.2M Na tartrate	**63.0**	0.1M TRIS pH8.5	**63.0**	15% PEG 8K
64.0	None	**64.0**	0.1M TRIS pH8.5	**64.0**	1M Na tartrate
65.0	0.2M Li_2 SO_4	**65.0**	0.1M TRIS pH8.5	**65.0**	30% MPD
66.0	None	**66.0**	0.1M TRIS pH8.5	**66.0**	3.0M $(NH_4)_2$ SO_4
67.0	0.05M $CaCl_2$	**67.0**	0.1M TRIS pH8.5	**67.0**	5% PEG 8K

68.0	None	**68.0**	0.1M TRIS pH8.5	**68.0**	3M Na/K phosphate
69.0	0.5M $MgCl_2$	**69.0**	0.1M TRIS pH8.5	**69.0**	10% PEG 8K
70.0	0.5M $(NH_4)_2 SO_4$	**70.0**	0.1M TRIS pH8.5	**70.0**	30% MPD
71.0	0.1M Li_2SO_4	**71.0**	0.1M Acetate pH4.6	**71.0**	10% Me2K PEG
72.0	0.2M $MgCl_2$	**72.0**	0.1M Acetate pH4.6	**72.0**	15% Me2K PEG
73.0	0.2M $(NH_4)_2 SO_4$	**73.0**	0.1M Acetate pH4.6	**73.0**	20% Me2K PEG
74.0	None	**74.0**	0.1M Acetate pH4.6	**74.0**	30% Me2K PEG
75.0	0.1M Li_2SO_4	**75.0**	0.1M Acetate pH4.6	**75.0**	10% PEG 4K
76.0	0.1M $MgCl_2$	**76.0**	0.1M Acetate pH4.6	**76.0**	15% PEG 4K
77.0	0.2M $(NH_4)_2 SO_4$	**77.0**	0.1M Acetate pH4.6	**77.0**	20% PEG 4K
78.0	None	**78.0**	0.1M Acetate pH4.6	**78.0**	30% PEG 4K
79.0	0.1M Li_2SO_4	**79.0**	0.1M Citrate pH5.6	**79.0**	10% Me2K PEG
80.0	0.2M $MgCl_2$	**80.0**	0.1M Citrate pH5.6	**80.0**	15% Me2KPEG
81.0	0.2M $(NH_4)_2 SO_4$	**81.0**	0.1M Citrate pH5.6	**81.0**	20% Me2k PEG
82.0	None	**82.0**	0.1M Citrate pH5.6	**82.0**	30% Me2K PEG
83.0	0.1M Li_2SO_4	**83.0**	0.1M Citrate pH5.6	**83.0**	10% PEG 4K
84.0	0.2M $MgCl_2$	**84.0**	0.1M Citrate pH5.6	**84.0**	15% PEG 4K
85.0	0.2M $(NH_4)_2 SO_4$	**85.0**	0.1M Citrate pH5.6	**85.0**	20% PEG 4K
86.0	None	**86.0**	0.1M Citrate pH5.6	**86.0**	30% PEG 4K
87.0	0.1M Li_2SO_4	**87.0**	0.1M MOPS pH6..5	**87.0**	10% Me2K PEG
88.0	0.2M $MgCl_2$	**88.0**	0.1M MOPS pH6..5	**88.0**	15% Me2K PEG
89.0	0.2M $(NH_4)_2 SO_4$	**89.0**	0.1M MOPS pH6..5	**89.0**	20% Me2K PEG
90.0	None	**90.0**	0.1M MOPS pH6..5	**90.0**	30% Me2K PEG
91.0	0.1M Li_2SO_4	**91.0**	0.1M MOPS pH6..5	**91.0**	10% PEG 4K
92.0	0.2M $MgCl_2$	**92.0**	0.1M MOPS pH6..5	**92.0**	15% PEG 4K
93.0	0.2M $(NH_4)_2 SO_4$	**93.0**	0.1M MOPS pH6..5	**93.0**	20% PEG 4K
94.0	None	**94.0**	0.1M MOPS pH6..5	**94.0**	30% PEG 4K
95.0	0.2M $Li_2 SO_4$	**95.0**	0.1M HEPES pH7.5	**95.0**	10% Me2K PEG
96.0	0.2M $MgCl_2$	**96.0**	0.1M HEPES pH7.5	**96.0**	15% Me2K PEG
97.0	0.2M $(NH_4)_2 SO_4$	**97.0**	0.1M HEPES pH7.5	**97.0**	20% Me2K PEG
98.0	None	**98.0**	0.1M HEPES pH7.5	**98.0**	30% Me2K PEG
99.0	0.1M $Li_2 SO_4$	**99.0**	0.1M HEPES pH7.5	**99.0**	10% PEG 4K
100.0	0.2M $MgCl_2$	**90.0**	0.1M HEPES pH7.5	**90.0**	15% PEG 4K
101.0	0.2M $(NH_4)_2 SO_4$	**101.0**	0.1M HEPES pH7.5	**101.0**	20% PEG 4K
102.0	None	**102.0**	0.1M HEPES pH7.5	**102.0**	25% PEG 4K
103.0	0.1M Li_2SO_4	**103.0**	0.1M TRIS pH8.5	**103.0**	10% Me2K PEG
104.0	0.2M $MgCl_2$	**104.0**	0.1M TRIS pH8.5	**104.0**	15% Me2K PEG
105.0	0.2M $(NH_4)_2 SO_4$	**105.0**	0.1M TRIS pH8.5	**105.0**	20% Me2K PEG
106.0	None	**106.0**	0.1M TRIS pH8.5	**106.0**	30% Me2K PEG
107.0	0.2M Li_2SO_4	**107.0**	0.1M TRIS pH8.5	**107.0**	10% PEG 4K
108.0	0.2M $MgCl_2$	**108.0**	0.1M TRIS pH8.5	**108.0**	15% PEG 4K
109.0	0.2M $(NH_4)_2 SO_4$	**109.0**	0.1M TRIS pH8.5	**109.0**	20% PEG 4K
110.0	None	**110.0**	0.1M TRIS pH8.5	**110.0**	30% PEG 4K
111.0	0.2M $MgCl_2$	**111.0**	0.1M TRIS pH8.5	**111.0**	30% PEG 4K
112.0	0.2M $MgCl_2$	**112.0**	0.1M MOPS pH6.5	**1120**	20% PEG 8K

113.0	0.2M $CaCl_2$	**113.0**	0.1M TRIS pH8.5	**113.0**	30% PEG 4K
114.0	0.2M $CaCl_2$	**114.0**	0.1M MOPS pH6.5	**114.0**	20% PEG 8K

Buffer pH is that of a 1.0M stock prior to dilution with the other reagent components (MES is made as a 0.5M stock).

Appendix II

Additive ScreenI

1. 0.1 M Barium chloride (Cation)
2. 0.1 M Cadmium chloride (Cation)
3. 0.1 M Calcium chloride (Cation)
4. 0.1 M Cobalt chloride (Cation)
5. 0.1 M Copper chloride (Cation)
6. 0.1 M Magnesium chloride (Cation)
7. 0.1 M Manganese chloride (Cation)
8. 0.1 M Strontium chloride (Cation)
9. 0.1 M Yterbium chloride (Cation)
10. 0.1 M Zinc chloride (Cation)
11. 30% Ethylene glycol (Organic)
12. 30% Glycerol (Organic)
13. 30% 1,6 Hexanediol (Organic)
14. 30% MPD (Organic)
15. 50% PEG 400 (Organic)
16. 0.1 M Trimethylamine HCl (Chaotrope)
17. 1.0 M Guanidine HCl (Chaotrope)
18. 0.1 M Urea (Chaotrope)
19. 15% 1,2,3 Heptanetriol (Amphiphile)
20. 20% Benzamidine HCl (Amphiphile)
21. 30% Dioxane (Organic)
22. 30% Ethanol (Organic)
23. 30% Isopropanol (Organic)
24. 30% Methanol (Organic)

Additive Screen II

1. 1.0 M Sodium iodide (Ion)
2. 0.1 M L-cysteine (Reducing Agent)
3. 0.1 M EDTA, sodium salt (Chelator)
4. 0.1 M b-Nicotinamide adenine dinucleotide (Co-factor)
5. 0.1 M Adenosine-5Õ-triphosphate disodium salt (C-factor)
6. 30% w/v D(+)-Glucose (Carbohydrate)
7. 30% w/v D(+)-Sucrose (Carbohydrate)
8. 30% w/v Xylitol (Carbohydrate)
9. 0.1 M Spermidine (Polyamine)
10. 0.1 M Spermine-tetrahydrochloride (Polyamine)
11. 30% w/v 6-Aminocaproic acid (Linker)
12. 30% w/v 1,5-Diaminopentane dihydrochloride (Linker)
13. 30% v/v 1,6-Diaminohexane (Linker)
14. 30% w/v 1,8-Diaminooctane (Linker)
15. 1.0 M Glycine (Linker)
16. 0.3 M Glycyl-glycyl-glycine (Linker)
17. 0.1 M Hexaminecobalt trichloride (Polyamine)
18. 0.1 M Taurine (Linker)
19. 0.1 M Betaine monohydrate (Linker)
20. 5% w/v Polyvinylpyrrolidone K15 (Polymer)
21. 3.0 M Non-detergent sulfo-betaine 195 (Solubilizing Agent)

22. 2.0 M Non-detergent sulfo-betaine 201 (Solubilizing Agent)
23. 0.1 M Phenol (Dissociating Agent)
24. 30% v/v Dimethyl sulfoxide (Chaotrope)

Detergent Screen

1. 1.0 M Ammonium sulfate (Salt)
2. 1.0 M Cesium chloride (Salt)
3. 1.0 M Potassium chloride (Salt)
4. 1.0 M Lithium chloride (Salt)
5. 2.0 M Sodium chloride (Salt)
6. 0.5 M Sodium fluoride (Salt)
7. 2.0 M Sodium thiocyanate (Salt)
8. 30% w/v Dextran sulfate, sodium salt (Polymer)
9. 50% v/v Jeffamine M-600 (Organic, non-volatile)
10. 40% v/v 2, 5 Hexanediol (Organic, non-volatile)
11. 40% v/v 1,3 Butanediol (Organic, non-volatile)
12. 40% v/v Polypropylene glycol P400 (Organic, non-volatile)
13. 40% v/v 1,4 Butanediol (Organic, non-volatile)
14. 40% v/v tert-Butanol (Organic, volatile)
15. 40% v/v 1,3 Propanediol (Organic, volatile)
16. 40% v/v Acetonitrile (Organic, volatile)
17. 40% v/v Gamma butyrolactone (Organic, volatile)
18. 40% v/v n-Propanol (Organic, volatile)
19. 5% v/v Ethyl acetate (Organic, volatile)
20. 40% v/v Acetone (Organic, volatile)
21. 2.5% v/v Dichloromethane (Organic, volatile)
22. 7% v/v n-Butanol (Organic, volatile)
23. 40% v/v 2,2,2-Trifluoroethanol (Organic, volatile)
24. 0.1 M 1,4-Dithio-DL-threitol (Reducing agent)

Bibliography

Akhtar, M., Botting, N. P., Cohen, M. A. & Gani, D. (1987) Enantiospecific synthesis of 3-substituted aspartic acids via enzymic amination of substituted fumaric acids, *Tetra.*, **43**, 5899-5908.

Ambrosina, R., Baronem, G., Castronuovo, G., Ceccarini, C., Cultrera, O. & Elia, V. (1987) Protein ligand interaction. A calorimetric study of the interaction of oligosaccharides and hen ovalbumin glycopeptides with concanavalin A, *Biochemistry,* **26**, 3971-3975.

Andersen, P., Andersen, A. B., Sorenson, A. L. and Nagai, S. (1995) Recall of long-lived immunity to mycobacterium-tuberculosis infection in mice, *J. Immun.*, **154**, 3359-3372.

Archer, C. H. & Gani, D. (1993) Kinetics and mechanism of syn-elimination of ammonia from (2S, 3R)-3-methylaspartic acid by methylaspartase, *Chem. Comm.*, **2**, 140-142.

Ashton, P. R., Boyd, S. E., Brown, C. L., Jayaraman, N. & Stoddart, J. F. (1997) A convergent synthesis of a carbohydrate-containing dendrimer, *Angew. Chem. Int. Ed. Engl.*, **37**, 732-735.

Baldock, C., Rafferty, J. B., Sedelnikova, S. E., Baker, P. J., Stuitje, A. R., Slabas, A. R., Hawkes, T. R. and Rice, D. W. (1996) A mechanism of drug action revealed by structural studies of Enoyl Reductase, *Science*, **274**, 2107-2110.

Baldock, C., de Boer, G-J., Rafferty, J. B., Stuitje, A. R. Rice D. W. (1998) Mechanism of action of Diazoboranes, *Biochem. Pharmacol.*, **55**, 1541-1549.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., Collins, D, Delisle, G. and Jacobs, W. R. (1994) *inh*A, a gene encoding a target for isoniazid and ethionamide in mycobacterium-tuberculosis, *Science,* **263**, 227-230.

Barker, A. H., Smyth, D. R., Marilyn-Wilson, R. & Weissbach, H. (1958) The purification and properties of β-Methylaspartase, *J. Biol. Chem.*, **234**, 320-328.

Belanger, A. E., Besra, G. S., Ford, M. E., Mikusova, K., Belisle, J. T., Brennan, P. J. and Inamine, J. M. (1996) The *embAB* genes of *Mycobacterium avium* encode an arabinosyl transferase involved in the cell wall arabinan biosynthesis that is the target for the antimycobacterial dug ethambutol, *P. N. A. S.*, **93**, 11919-11924.

Bellamacina, C. (1996) The nicotinamide dinucleotide binding site motif: a comparison of nucleotide binding proteins, *Protein Motifs*, **10**, 1257-1269.

Belisle, J. T., Vissa, V. D., Mikusova, K., Exekowitz, R. A. B., Joiner, K. A. (1997) Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis, *Science*, **276**, 1420-1422.

Besra, G., Morehouse, C. B., Rittner, C. M., Waechters, C. J. and Brennan, P. J. (1997) Biosynthesis of mycobacterial lipoarabinomannan, *J. Biol. Chem.*, **272**, 18460-18466.

Bhattacharyya, L., Ceccarini, C., Lorenzoni, P. and Brewer, C.F. (1987) Concanavalin-A interactions with asparagine-linked glycopeptides - bivalency of high mannose and bisected hybrid type glycopeptides, *J. Biol. Chem*, **262**, 1288- 1293.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ColladoVides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden,M. A., Rose, D. J., Mau, B. and Shao. Y. 91997) The complete genome sequence of Escherichia coli K-12 ,*Science*, **277**, 1453 – 1470

Bloom, B. R. & Fine, P. E. M. in "Tuberculosis: pathogenesis, protection and control", (Ed. B. R. Bloom), *Washington: ASM Press*, 1994, 53-57.

Boehringer Mannheim UK. (1993) Diagnostics and Biochemicals Ltd, Bell Lane, Lewes, East Sussex, BN7 1JG.

Botting, N. P., Jackson. A. A. & Gani, D. (1989) 15N-Isoptope and double fractionation studies of the mechanism of 3-methylaspartase: Concerted elimination of ammonia from (2S, 3S)-3-methylapspartic acid, *J. Chem. Soc. Chem. Comm.*, 1583-1585.

Botting, N. P. & Gani, D. (1992) Mechanism of C-3 hydrogen exchange and the elimination of ammonia in the 3-methylasparatate ammonia-lyase reaction, *Biochem*istry, **31**, 1509-1520.

Bourne,Y., Ayouba, A., Rouge, P. and Cambillau, C. (1994) Interaction of a legume lectin with 2 components of the bacterial-cell wall – A crystallographic study, *J. Biol. Chem.*, **269**, 9429 – 9435.

Bourne, Y., Bolgiano. B., Liao, D. L., Strecker, G., Cantau, P., Herzberg, O., Feizi, T. and Cambillau, C. (1994) Cross-linking of mammalian lectin (galectin-1) by complex biantennary saccharides, *Nat. Struc. Biol.*, **12**, 863-870.

Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilising the principle of protein dye binding. *Anal. Biochemistry*, **72**, 248-254.

Brazeau, C., Gottschalk, M, Vincelette, S and Martineau- Doize, B. (1996) In vitro phagocytosis and survival of Streptococcus suis capsular type 2 inside murine macrophages, *Microbiology*, **142**, 1231-1237.

Brennan, P. J. and Ballou, C. E. (1967) The biosynthesis of mannophosphoinositides by *Mycobacterium phlei*, *J. Biol. Chem.*, **242**, 3046-3056.

Brunger, A.T., Adams, P. D, Clore, G. M, DeLano. W.L., Gros, P, GrosseKunstleve, R.W., Jiang, J.S., Kuszewski, J.Nilges, M., Pannu, N. S., Read, R.J., Rice, L. M., Simonson, T. and

Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Cryst.* **D54**, 905-921.

Cole, S. T., Brosch. R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles_S, Hamlin, H., Holroyd, S., Hornby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence, *Nature*, *393*, 537-544.

Chatterjee, D. (1997) The mycobacterial cell wall: structure, biosynthesis and site of drug action. *Cur. Op. In Chem. Biol.*, **1**, 579-588.

Chayen, N. E., Stewart, P. D. S. and Blow, D. M. (1992) Microbatch crystallisation under oil - a new technique allowing many small-volume crystallisation trials *J. Cryst. Growth*, **122**, 176-180.

Chervenak, M. C., Oas, T. G. and Toone, E. J. (1992) Structure and energetics of concanavalin A oligosaccharide binding, *Abs. Am. Chem. Soc.*, **204**, Pt1, 102-CARB.

Chervenak, M. C. and Toone, E. J. (1994) A direct measurement of the contribution of solvent reorganisation to the enthalpy of ligand-binding, *J. Am. Chem. Soc.*, **116**, 10533-10539.

Chervenak, M. C. and Toone, E. J. (1995) Calorimetric analysis of the binding specificities of oligomannoside-binding proteins using methylated monosaccharides, *Bioorg. Med. Chem.*, **4**, 1963-1973.

Chung, K. N., Walter, P., Aponte, G. W. and Moore, H. P. H. (1989) Molecular sorting in the secretory pathway, *Science*, **243,** 192-197.

Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography, *Acta Cryst.* **D50**, 760-763.

Cudney, B., Patel, S., Weisgraber, K. and Newhouse, Y. (1994) Screening and optimisation strategies for macromolecular crystal growth, *Acta Cryst.,* **D50**, 414-423.

Cunningham, A.F. and Spreadbury, C. L. (1998) Mycobacterial stationary phase induced by low oxygen tension: Cell wall thickening and localization of the 16-kilodalton alpha-crystallin homology, *J. Bacteriology*, **180**, 801-808.

Daffe, M., Brennan, P. J. and McNeil, M. (1990) Predominant structural features of the cell wall arabinogalactan of Mycobacterium tuberculosis as revealed through characterisation of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by 1H and 13C NMR analysis, *J. Biol. Chem.*, **265,** 6734-6743.

Danielson, S. J. and Gray, G. R. (1982) Structures of the 2 homologous series of dialkene mycolic acids from mycobacterium-smegmatis, *J. Biol. Chem.*, **257**, 2196- 2203.

D'Arcy, A. (1994) Crystallising proteins - A rational approach, *Acta Cryst.*, **D50**, 469-471.

D'Arcy, A., Elmore, C., Stihle, M. and Johnston, J.E. (1996) A novel approach to crystallising proteins under oil, *J. Cryst. Growth*, **168**, 175-180.

De Boer, G.-J., Pielage, G. J. A., Nijkamp, J. J.,Slabas, A. R., Rafferty, J. B., Baldock, C., Rice, D. W. and Stuitje, A. R. (1999) Molecular and genetic analysis of enoyl-acyl carrier protein reductase inhibition by diazoborine, *Mol. Micro.*, **31**, 443-450.

Deacon, A., Gleichmann, T.. Kalb, J., Price, H., Rafferty, J., Bradbrook, G., Yariv, J. and Helliwell, J. R. (1997) The structure of concanavalin A and its bound solvent structure determined with small molecule accuracy at 0.94Å resolution. *J. Chem. Soc., Faraday Trans.*, **53**, 4305-4312.

Deng, L., Mikusova, K., Robuck, K. G., Schermann. M., Brennan. P. J. and McNeil, M. R. (1995) Recognition of the multiple effects of ethambutol on the metabolism of the mycobacterial cell envelope, *Antimicrob. Agents Chemother.*, **39**, 694-701.

Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb (Gilboa), A. J., Dodsen, E., Paiz, M., Wan, T. and Campbell, J. W. (1989) The structure of the saccharide-binding site of concanavalin-A, *EMBO J.*, **8**, 2189-2193.

Dessen, A., Quemard, A., Blanchard, J. S., Jacobs, W. R. and Sacchettini, J. C. (1995) Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis, *Science*, **267**, 1638-1641.

Dolin, P., Dye, C., Raviglione, M. and Kochi, A. (1998) Current status of the global tuberculosis epidemic, *Nature*. http://www.nature.com/Nature2/serve?SID=25602728

Doublie, S. (1997) Preparation of selenomethionyl proteins for phase determination *Meth. Enzymol.*, **276**, 523-530.

Draper, P. (1992) in The Biology of Mycobacteria, Vol 1, (Ratledge, C. C. and Stanford, J. eds) pp9-52, Academic Press. London.

Draper, P., Khoo, K. H., Chatterjee, D, Dell, A. and Morris, H. R. (1997) Galactosamine in walls of slow growing mycobacteria, *J. Biochem.*, *327*, 519-525.

Ducruix, A and Giege, R. (1992) in Crystallisation of nucleic acids and proteins: A practical approach, Ducruix, A and Giege, R. eds. IRL Press Ltd, Oxford, pp 73-97.

Duncan, K. (1997) Towards the next generation of drugs and vaccines for tuberculosis, *Chem. Indus.*, **21**, 861-865.

Dwek, R.A. (1996) Glycobiology: Toward understanding the function of sugars, *Chem. Revs.*, **96**, 683-720.

Engh, R. A and Huber, R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement, *Acta. Cryst.* **A47**, 392-400.

Elliott, T., Hastings, M. and Desselberger, U. (1996) in Medical Microbiology. 3rd Edition, Blackwell Sciences, pp51-55.

Etemadi, A., Farid, R. and Stanford, J. L. (1992) Immunotherapy for drug resistant tuberculosis, *Lancet*, **340**, 1360-1361.

Evans, J. (1998) TB: know your enemy, *Chem. Britain*, **11,** 38-42

Evans, P. (1993) "Data Reduction", Data Collection and Processing, Proceedings of the CCP4 Study Weekend on Data Collection and Processing, 114-122.

Ferre-D'Amare, A. R. and Burley, S. K. (1996) Dynamic light scattering in evaluating crystallisability of macromolecules, *Meth. Enzymol.*, **276**, 157-166.

Frieden, T. R., Sherman, L. F., Maw, K. L., Fujiwara, P. I., Crawford, J. T., Nivin, B, Sharp, V., Hewlett, D., Brudney, K., Alland, D. and Kreiswirth, B. N. (1996) A multi-institutional outbreak of highly drug-resistant tuberculosis - Epidemiology and clinical outcomes, *J. Amer. Medic. Assoc.*, **276**, 1229-1235.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., Mckenney K., Sutton, G., Fitzhugh, W., Fields, C. A., Gocayne, J. D., Scott, J. D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J., F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., Mcdonald, L.A., Small K. V., Fraser, C. M., Smith, H. O. and Venter, J. C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae*, *Science*, **269**, 496-512.

George, K. M., Yuan, Y., Sherman, D. R. and Barry, C. E. (1995) The biosynthesis of cycloprotonated mycolic acids in Mycobacterium tuberculosis – identification and functional analysis of CMAS-2, *J. Biol. Chem.*, **270**, 27292-27298.

Gewirth, D. (1996) "An oscillation data processing suite for macromolecular crystallography", The HKL Manuel, 96-97.

Goda, S., Minton, N. P., Botting, N. P. & Gani, D. (1992) Cloning, sequencing and expression in *Escherichia coli* of the clostridium tetanomorphum gene encoding β-methylaspartase and characterisation of the recombinant protein, *Biochemistry*, **31**, 10747-10756.

Goldstein, I. J. and Poretz, R. D. (1986) Isolation and chemical properties of lectins. The lectins: Properties, Functions and Applications. In Biology and Medicine. I. E. Leiner, N. Sharon and I. J. Goldstein. eds. Orlando, Academic Press Inc., pp. 35-244.

Gottscalk, M., Kolberg, J., Charland, N and Jacques, M. (1995) *Streptococcus pneumonia* types 19A and 19F and *Streptococcus suis* capsular type 8 share common capsular epitopes, *J. Clin. Microbiology*, **33**, 2492-2495.

Grassberger, M. A., Turnowsky, F. and Hildebrandt, J. (1984) Preparation and antibacterial activities of new 1,2,3 - diazaborine derivatives and analogs, *J. Medic. Chem.*, **27**, 947-953.

Grunwald, E. and Steel, C. (1995) Solvent reorganisation and thermodynamic enthalpy entropy compensation, *J. Am. Chem. Soc.*, **117**, 5687-5692.

Gunther, G. R., Wang, J. L., Yahara, I., Cunningham, B. A. and Edelman, G. (1973) Concanavalin A derivatives with altered biological activities, *P.N.A.S.*, **70**, 1012-1016.

Gupta, D., Bhattacharyya, L., Fant, J., Macaluso, F., Sabesan, S, and Brewer, C.F. (1994) Observation of unique cross-linked lattices between multiantennary carbohydrates and soybean lectin - presence of pseudo- 2-fold axes of symmetry in complex type, *Biochemistry*, **33**, 7495-7504.

Hampton research, crystallisation research tools, info and ideas. 27632 El Lazo Road, Suite 100, Laguna Niguel, CA 92677-3913. (1995) 5, 1, 17.

Ingrosso, D, Fowler, A. V, Bleibaum, J. and Clarke, S. (1989) Sequence of the D-aspartyl L-isoaspartyl protein methyltransferase from human-erythrocytes – common sequence motifs for protein, DNA, RNA, and small molecule s-adenosylmethionine - dependent methyltransferases, *J. Biol. Chem.*, **264**, 20131- 20139.

Janknecht, R., Demartynoff, G., Lou, J., Hipskind, R.A., Nordheim, A. and Stunnenberg, H. G. (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus, *P.N.A.S.*, **88**, 8972-8976.

Jencks, W.P. (1981) On the attribution and additivity of binding-energies, *P. N. A. S.* **78**, 4046- 4050.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Cryst.*, **A47**, 110-119.

Kazlauskas, R. J. (1994) Elucidating structure-mechanism relationships in lipases - prospects for predicting and engineering catalytic properties, *Trends Biotechnol.*, **12**,464-472

Koch, R (1882) *Berliner.Wochenschr.* **19**, 221-230.

Koplin, R., Brisson, J. and Whitfield, C. (1997) UDP-galactofuranose precursor required for formation of the lipopolysaccharide O antigen of *Klebsiella pneumoniae* serotype 01 is synthesised by the product of the rfbd (KP01) gene *J. Biol. Chem.*, **272**, 4121-4128.

Kornfeld R. and Kornfeld, S. (1985) Assembly of asparagine linked oligosaccharides. *Annu. Rev. Biochem.*, **54**, 631-664.

Laskowski, R. A., MacArthur, M W., Moss, D. S. and Thorton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, J. *Appl. Cryst.*, **26**, 548-558.

Latge, J-P., Kobayashi, H., Debeaupuis, J-P., Diaquin, M., Sarfati, J., Wieruszeski, J-M., Parra, E., Bouchara, J-P and Fournet, B. (1994) Chemical and immunological chareterisation of the extracellular galactomannanof *Aspergillus fumigatis. Infect. Immun.*, **62**, 5425-5433.

Lee, R. E., Mikusova, K., Brennan, P. J. and Besra, G. S. (1995) Synthesis of the mycobacterial arabinose donor β-D- arabinofuranosyl-1-monophosphoryldecaprenol, development of a basic arabinosyl-transferase assay, and identification of ethambutol as an arabinosyl transferase inhibitor, *J. Am. Chem. Soc.*, **117**, 11829-11832.

Leslie, A. G. W. (1993) "Autoindexing of rotation diffraction images and parameter refinement", Data Collection and Processing, Proceedings of the CCP4 Study Weekend on Data Collection and Processing, 44-51.

Liu, J., Rosenberg, E. Y. and Nikaido, H. (1995) Fluidity of the lipid domain of cell-wall from mycobacterium- chelonae, *P. N. A. S.*, **92**, 11254-11258.

Lowe, J.B. and Ward, P.A. (1997) Therapeutic inhibition of carbohydrate-protein interactions *in vivo*, *J. Clin. Invest.*, **99**, 822- 826.

Ma, Y, Mills, J., Belisle, J., Vissa, V, Howell, M., Bowlin, K., Scherman, M. S. and McNeil, M. (1997) Determination of the pathway for rhamnose biosynthesis in mycobacteria : cloning, sequencing and expression of the gene encoding α-D-glucose-1-phosphate thymidylyltransferase, *Microbiology*, **143**, 937-945.

Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. and Stover, C. K. (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis, *J. Bacteriol.*, **178**, 1274-1282.

Mandal D.K. and Brewer, C.F. (1992) Cross-linking activity of the 14-kilodalton beta-galactoside- specific vertebrate lectin with asialofetuin - comparison with several galactose-specific plant-lectins, *Biochemistry*, **31**, 8465-8472.

Mandal, K. D. and Brewer, F. C. (1993) Interactions of concanavalin A with glycoproteins: Formation of homogenous glycoprotein-lectin cross-linked complexes in mixed precipitation systems, *Biochemistry*, **31**, 12602-12609.

Mandal, K. D., Kishore, N. and Brewer, F. C. (1994) Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A, *Biochemistry*, **33**, 1149-1156.

Maniatis. T., Fritsch. E. F. & Sambrook. J. (1982). *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor. NY: Cold Spring Harbor Laboratory.

Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol.*, **33**, 491-497.

McAdam, R. A., Weisbrod, T.R., Martin, J., Scudero, J. D., Brown, A. M., Cirillo, J. D., Bloom, B. R. and Jacobs, W. R. (1995) In-vivo growth-characteristics of leucine and methionine auxotrophic mutants of mycobacterium-bovis BCG generated by transposon mutagenesis, *Infect. Immun.*, **63**, 1004-1012.

McNeil, M. R., Daffe, M. and Brennan, P. J. (1990) Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell wall, *J. Biol. Chem.*, **265**, 18200-18206.

Meunier, S. J. & Roy, R. (1996) Polysialosides scaffolded on p-tert-butylcalix[4]arene, *Tetrahedron Lett.*, **37**, 5469-5472.

Middlebrook, G. (1952) *Am. Rev. Tuberc. Pulm. Dis.*, **65**, 765.

Mikusova, K., Mikus, M., Besra, G., Hancock, I. and Brennan, P. J. (1996) Biosynthesis of the linkage region of the mycobacterial cell wall, *J. Biol. Chem.*, **271**, 7820-7828.

Miller, T.M., Kwock, E.W. and Neenan, T.X. (1992) Synthesis of 4 generations of monodisperse aryl ester dendrimers based on 1,3,5-benzenetricarboxylic acid, *Macromol.*, **25**, 3143-3148.

Minnikin, D.E. (1982) The biology of the mycobacteria, London: Academic Press, 95-184.

Mikol, V., Hirsch. and Giege, R. (1990) Diagnostic of precipitant for biomacromolecule crystallization by quasi-elastic light-scattering *J. Mol. Biol.*, 213, 187-195.

Mikusova, K., Slayden, D. A., Besra, G.S. and Brennan, P.J. (1995) Biogenesis of the mycobacterial cell wall and the site of action of ethambutol, *Antimicrob. Agents Chemother.* **39**, 2482-2489.

Moothoo, D. N. and Naismith J. H. (1998a) A general method for the co-crystallisation of concanavalin A with carbohydrates, *Acta Cryst*, **D54**, 1-3.

Moothoo, D. N. and Naismith J. H. (1998b) Concanavalin A distorts the β-GlcNAc-(1-2)-Man linkage of β-GlcNAc-(1-2)-α-Man-(1-3)-[β-GlcNAc-(1-2)-α-Man(1-6)]-Man upon binding, *Glycobiology*, **8**, 173-181.

Moraes, C. T., Bosch, M. and Parodi, A. J. (1987) Structural characterisation of several galactofuranose containing high mannose type oligosaccharides present in glycoproteins of the trypanosomatid Leptomonas samueli, *Biochem.*, **27**, 1543-1549.

Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb (Gilboa), A. J. and Yariv, J. (1994) Refined structure of concanavalin-A complexed with methyl alpha-d- mannopyranoside at 2.0 angstrom resolution and comparison with the saccharide free structure, *Acta. Cryst.* **D 50**, 847-858.

Naismith, J. H. & Field, R. A. (1996) Structural basis of trimannoside recognition by concanavalin-A, *J. Biol. Chem.*, **271**, 2, 972-976.

Nassau, P. M, Martin, S. L., Brown, R. E., Weston, A., Monsey, D., McNeil, M R. and Duncan, K. (1996) Galactofuranose biosynthesis in Escherichia coli K-12: identification and cloning of UDP-Galactopyranose mutase, *J. Bacteriology*, **178**, 4, 1047-1052.

Navaza, J. (1994) AMoRe: An automated package for molecular replacement, *Acta Cryst.* **A50**, 157-163.

Otwinowski, Z & Minor, W. (1996) Processing of X-ray diffraction data collected in oscillation mode, *Meth. Enzym.*, **276**, 307-326.

Page, D., Aravind, S. & Roy, R. (1996) Synthesis and lectin binding properties of dendritic mannopyranoside, *J. Chem. Soc. Chem. Comm.*, 1913-1914.

Pollard, J., Richardson, S., Lasry, P., Archer, C. H., Neal, T., Akhtar, M., Botting, N. P. & Gani, D. (1998) Mechanism of 3-methylaspartase probed using 2H- and 15N-isotope effects and active site directed reagents: A flip from a concerted to a carbocationic elimination upon changing the C-3 stereochemistry in the substrate from R to S, *In Press*.

Peake, P., Gooley, A. and Britton, W. J. (1993) Mechanism of interaction of the 85b secreted protein of *Mycobacterium bovis* with fibronectin, *Infect. Immun.*, **61**, 4828-4834.

Potera, C. (1998) A sweet way to keep proteins safe, *Science*, **281**, 1793.

Quemard, A., Lacave, C. and Laneelle, G. (1991) Isoniazid inhibition of mycolic acid synthesis by cell-extracts of sensitive and resistant strains of mycobacterium-aurum, *Antimicrob. Agents Chemother.*, **35**, 1035-1039

Quemard, A., Sachhettini, J. C., Desssen, A., Vicheze, C., Bittman, R., Jacobs Jr., W. R. and Blanchard, J. S. (1995) Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*, *Biochemistry*, 34, 8235-8241.

Quesenberry, M. S., Lee, R. T. and Lee, Y. C. (1997) Difference in the binding mode of two mannose-binding proteins: Demonstration of a selective minicluster effect, *Biochemistry*, 36, 2724-2732.

Quiocho, F. A. (1993) Probing the Atomic Interactions Between Proteins and Carbohydrates, *Biochem. Soc. Trans.*, **21**, 442-448.

Raetz, C. R. H., (1990) Biosynthesis of endotoxins. *Annu. Rev. Biochem.* **59**, 129-170.

Ramakrishnan, C. and Ramachandran, G. N. (1965) Stereochemical criteria for polypeptide and protein chain conformation, *J. Biophys.*, **5**, 909-933.

Rini, J. M. (1995) Lectin Structure, *Annu. Rev. Biophys. Biomol. Struc.*, **24**, 551-577.

Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L and Carver, J P. (1993) X-ray crystal structure of a pea lectin-trimannoside complex at 2.6Å resolution, *J. Biol. Chem.*, **268**, 10126-101132.

Reeke, G. N., Jr., Becker, J., and Edelman, G. M. (1975) The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure, *J. Biol. Chem.***250**, 1525-1547.

Rodgers, D. W. (1997) Practical Cryocrystallography, *Meth. Enzym.*, **276**, 183-203.

Rossmann, M. G., Liljas, A., Branden, C. I. And Banaszak, L. J. (1975) *The Enzymes, 3rd edition* (Boyer, P. D., ed) **11**, 61-102, Academic press, New York.

Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs Jr, W. R. and Sacchettini, J. C. (1998) Modification of the NADH of the isoniazid target (inhA) from *Mycobacterium tuberculosis*, *Science*, **279**, 98-102.

Salemme, F. R. (1985) Protein crystallisation by free interface diffusion, *Meth. Enzymol.*, **114**, 140-141.

Salo, W. L., Aufderheide, A. C., Buikstra, J. and Holcomb, T. A. (1994) Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvians mummy, *P. N. A. S.* **91**, 2091-1094.

Sairam, M. R. (1989) Role of carbohydrates in glycoprotein hormone signal transduction, *FASEB J.*, **3**, 1915-1926.

Sauer, O., Schmidt, A. and Kratky, C. (1997) Freeze-trapping isomorphous xenon derivatives of protein crystals, *J. Appl. Cryst.*, **30**, 476-486.

Scherman, M. S., Bournonville, L., K., Bush, D., Xin, Y., Deng, L. and McNeil, M. (1996) Polyprenylphosphate-pentoses in Mycobacteria are synthesised from 5-phosphoribose pyrophosphate, *J. Am. Chem. Soc.*, **271**, 29652-29658

Schiering, N., Kabsch, W., Moore, M.J., Distefano, M. D., Walsh, C. T. and Pai, E. F. (1991) Structure of the detoxification catalyst mercuric ion reductase from bacillus sp strain-rc607, *Nature*, **352**, 168-172.

Sharon, N. and Lis, H. (1989) Lectins as cell recognition molecules, Science, 246, 227-233.

Sigal, G. B., Mammen, M., Dahmann, G. & Whitesides, G. M. (1996) Polyacrylamides bearing pendant alpha-sialoside groups strongly inhibit agglutination of erythrocytes by influenza virus: the strong inhibition reflects enhanced binding through co-operative polyvalent interactions, *J. Am. Chem Soc.*, **118**, 3789-3800.

Singh, S. and Hogan, S. E. (1994) Isolation and characterisation of sugar nucleotides from *Mycobacterium smegmatis*, *Microbios.*, **77,** 217-222.

Staats, J. J., Feder I., Okwumabua, O and Chengappa, M. M. (1997) *Streptococcus suis*: Past and present, *Vet. Res. Comm.*, **21**, 381-407.

Stanford, J. L., Bahr, G. M., Rook, G. A. W., Shabaan, M. A., Chugh, T, D., Gabriel, M., Alshimali, B., Siddiqui, Z, Ghardani, F. and Behbehani, K. (1990) Immunotherapy with mycobacterium –vaccae an adjuct to chemotherapy in the treatment of pulmonary tuberculosis, *Tubercle*. **71**, 87-93.

Stevenson, G., Neal, D., Liu, M., Hobbs, N. H., Packer, M., Batley, J., Redmond, W., Lindquist, L and Reeves, P. (1994) Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster, *J. Bacteriol.*, **176**, 4144-4156.

Studier, F. W. and Moffat, B. A. (1986) Use of bacteriophage-T7 RNA polymerase to direct selective high level expression of cloned genes, *J. Mol. Biol.*, **189**, 113-130.

Stura, E. A. and Wilson, I. A. (1992) in Crystallisation of nucleic acids and proteins: A practical approach, Ducruix, A and Giege, R. eds. IRL Press Ltd, Oxford, pp 99-126.

Sumner, J. B. (1919) The Globulins of the Jack Bean Canavalia ensiformis, *J. Biol. Chem.*, **37**, 137-142.

Taillefumier, C., de Fornel, D. and Chapleur, Y. (1996) Design and synthesis of bisubstrate analoguesfor 3-hydroxy-3-methyl-glutaryl coenzyme A reductase, *Bioorg. Med. Chem. Lett.*, **6**, 615-618.

"TB a crossroads" in W. H. O. Report on the Global Tuberculosis Epidemic (1998)

Tayaro, J., Besnier, J. M., Laudat., Cattier, B. and Choutet. P. (1996) Infective Endocarditis due to *Streptococcus suis* serotype 2, *Eur. J. Clin. Microbiology Infect. Dis.*, **15**, 765-766.

Terwillinger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Cryst.*, **D55**, 849-861.

Thoden, J. B. and Holden, H. M. (1998) Dramatic differences in the binding of UDP-Galactose and UDP-Glucose-4-epimerase from *Escherichia coli*, *Biochemistry,* **37,** 11469-11477.

Thomas, J. P., Baughn, C. O., Wilkinson, R. G. and Shepard, G. (1961) Am. Rev. *Respir Dis.*, **83**, 891-893.

Toone, E. J. (1994) Structure and energetics of protein carbohydrate complexes, *Curr. Opin. Struct. Biol.*, **4**, 719-728.

Tsukioka, Y., Yoshihisa, Y., Nakano, Y., Oho, and Koga, T. (1997) Identification of a fourth gene involved in dTDP-rhamnose synthesis in Streptococcus mutans, *J. Bacterio.*, **179**, 4411-4414.

Tsukioka, Y., Yoshihisa, Y., Nakano, Y., Oho, and Koga, T. (1997) Biological function of the dTDP-rhamnose synthesis pathway in Streptococcus mutans, *J. Bacterio.*, **179**, 1126-1134.

Tuberculosis Treatment Observer *from the WHO.* (1998)

Ulmer, J. B., Donnelly, J. J.,Parker, S. E., Rhodes, G. H., Felgner, P. L., Dwarki, V. J., Gromkowski, S. H., Deck, R. R., Dewitt, C. M.. Friedman, A., Hawe, L. A., Leander, K. R., Martinez, D, Perry, H. C., Shiver, J. W., Montgomery, D. L. and Liu, M. A. (1993) Heterologous protection against influenza by injection of DNA encoding a viral protein, *Science*, **259**, 888-892.

Unkefer, C. J. and Gander, J. E. (1979) The 5-O-(-D-galactofuranosyl containing glycopeptide from Penicillium charlesii, *J. Biol. Chem.* **254**, 13131-12135.

Weis,W.I. and Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.*, **65**, 441-473.

Weston, A., Stern, R. J., Lee, R. L., Nassau, P. M., Monsey, D., Martin, S, L., Schermann, M. S., Besra, G. S., Duncan, K and McNeil, M. R. (1998) Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues, *Tuberc. Lung Dis.*, **78**, 123-131.

Whitfield, C. (9195) Biosynthesis of lipopolysaccharides O antigens, *Trends Microbiol.* **3**, 178-185.

Winkler, M. F. & Williams, V. R. (1967) New substrates for β-methylaspartase, *Bioch. Biophys. Acta.* **146**, 287-289.

Withers, S. G. (1999) 1998 Hoffmann la Roche Award Lecture - Understanding and exploiting glycosidases, *Canadian journal of chemistry-revue canadienne de chimie*, **77**, 1-11.

Woluka, B. A., McNeil, M. R., Hoffmann, E., Chojnacki, T. and Brennan, P. J. (1994) Recognition if the lipid intermediate for arabibogalactan/arabinomannan biosynthesis and its relation to the mode of action of Ethambutol on mycobacteria, *J. Biol Chem*, **269**, 23328-23335.

Wooten, E. W., Bazzo, R., Edge, C. J., Zamze, S., Dwek, R. A. and Rademacher, T. W. (1990) Primary sequence dependence of conformation in oligomannose oligosaccharides, *Euro. Biophys.*, **18**, 139-148

Wright, C. S. and Hester, G. (1996) The 2.0 angstrom structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose – evidence for 2 unique binding modes, *Structure*, **4**, 1339-1352.

Young, D. B. (1995) The return of an old nemesis: tuberculosis, *Odyssey*, **1**, 10-16.

Zulauf, M. and D'Arcy, A. (1992) Light-scattering of proteins as a criterion for crystallisation, *J. Cryst. Growth*, **122**, 102-106.

Zanini, D. & Roy, R. (1997) Chemoenzymatic synthesis and lectin binding properties of dendritic N-acetyllactosamine, *Bioconjugate, Chem.*, **8** 187-192.

Zhang, Y., Heym, B.. Allen, B., Young, D. and Cole, S. (1992) The catalase- peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis, *Science*, **358**, 591-593.

Zhong, S., Jordan, F., Kettner, C. and Polgar, L. (1991) Observation of tightly bound ^{11}B nuclear magnetic resonance signals on serine proteases. Direct solution evidence for tetrahedral geometry around the boron atom in the putative transition state analogues, *J. Amer. Chem. Soc.*, **113**, 9429-9435.

Publications

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Initiating a crystallographic study of UDP-galactopyranose mutase from *Escherichia*

Stephen A. McMahon,[a] Gordon A. Leonard,[b] Louise V. Buchanan,[a] Marie-France Giraud[a] and James H. Naismith[a]*

[a]Centre for Biomolecular Sciences, School of Biomedical Science, North Haugh, The University, St Andrews, Fife, KY16 9ST, Scotland, and [b]Joint Structural Biology Group, European Synchrotron Radiation Facility, BP220, F-38043 Grenoble CEDEX, France

Correspondence e-mail: naismith@st-and.ac.uk

Received 9 July 1998
Accepted 14 August

UDP-galactopyranose mutase, the enzyme responsible for the conversion of UDP-galactopyranose to UDP-galactofuranose, has been crystallized in a form suitable for X-ray diffraction studies. UDP-galactofuranose is a key component of mycobacterial cell walls. Crystals of both the native protein and a selenomethionine variant have been grown by the vapour-diffusion method in hanging drops, and diffract to beyond 3.0 Å using synchrotron radiation. Equilibration was against a solution of 20%(*w*/*v*) polyethylene glycol (4K), 12%(*v*/*v*) 2-propanol, 0.1 *M* HEPES pH 7.6 at 293.5 K. Crystals grow as thin plates of dimensions 0.4 × 0.2 × ~0.02 mm. They are orthorhombic, space group $P2_1$, with unit-cell dimensions $a = 71.12$, $b = 58.42$, $c = 96.38$ Å, $\beta = 96.38°$. 92% (native) and 94% (selenomethionine) complete data sets have been recorded to 2.9 Å ($R_{\rm merge}$ = 5.0%) and 3.0 Å ($R_{\rm merge}$ = 6.9%), respectively. The Matthews coefficient is 2.35 Å^3 Da^{-1} for a dimer in the asymmetric unit, the solvent content being 47%. Diffraction data have also been recorded on a putative platinum derivative to 3.5 Å.

1. Introduction

The alarming increase in the incidence of tuberculosis worldwide led to the World Health Organization (WHO) declaring a 'global emergency' in 1993. It estimates that three million people die each year from tuberculosis and that this figure may rise to four million when the devastating link between tuberculosis and human immunodeficiency virus (HIV) is considered (Duncan, 1997). Being co-infected with HIV and *Mycobacterium tuberculosis* increases the risk of contracting a clinically significant disease from a 10% lifetime risk to around 10% a year (Young, 1995). Although in most cases of tuberculosis the present chemotherapy available is effective, it is apparent that more and more strains resistant to current therapeutic agents are emerging.

The cell wall of mycobacteria is essential for cell growth and survival in the host. It confers on bacteria an extraordinary resistance to drugs, the host immune system, dehydration and acid treatment (Draper, 1982). Ethambutanol, one of the most successful front-line drugs currently used to treat tuberculosis, functions by inhibiting cell-wall assembly (Winford, 1982). The structure of the mycobacterial cell wall is elaborate, consisting of several unusual carbohydrate polymers (glycans). The arabinogalactan layer covalently attaches a highly impermeable layer of mycolic acids to a peptidoglycan layer and is thus a key component of the cell wall. Arabinogalactan consists in part of alternating β-1,5 and β-1,6 galactofuranosyl residues. Galactofuranose is found only in bacteria and its synthesis is thus an intellectually appealing target for inhibition. The biosynthesis of UDP-galactofuranose (Gal_f) takes place by the

direct conversion of UDP-galactopyranose by UDP-galactopyranose mutase (mutase; Nassau *et al.*, 1995). Highly homologous enzymes (40% sequence identity) are found in both *Escherichia coli* and *Klebsiella pneumonia* (Weston *et al.*, 1998; Fig. 1). Analysis of the primary sequences identifies a consensus $\beta\alpha\beta$ (Rossmann fold) motif consistent with nucleotide sugar-binding enzymes. All three enzymes are known to incorporate flavin adenine dinucleotide (FAD), and the enzyme from *K. pneumonia* has recently been shown to require nicotinamide adenine dinucleotide (phosphate) (NADP) sugar-binding enzymes for activity (Koplin *et al.*, 1997). As the reaction carried out does not involve the net transfer of electrons, the mechanistic role of these cofactors is puzzling. In a program to study cell-wall biosynthesis, we have crystallized the native and selenomethionine variant enzyme from *E. coli*. Our experiments suggest that the *E. coli* enzyme also incorporates NAD(P).

2. Protein purification

All chemicals were purchased from Sigma (Poole, UK), except polyethylene glycol which was obtained from Fluka (Gillingham, UK). The clone for UDP-galactopyranose mutase was generously donated by Glaxo Wellcome.

5 ml of an enriched growth medium, Terrific Broth (Maniatis *et al.*, 1982), containing ampicillin (200 μg ml^{-1}) was inoculated with a single colony of BL21(DE3)(pORF6) and incubated overnight at 310 K with shaking. This was used to inoculate 500 ml of similarly treated media. Incubation continued under the same conditions until an OD_{600} of 0.6 was reached. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 m*M* to initiate pro expression. At this point the temperature was lowered 303 K and incubation was continued for a further 3 h, a which the cells were harvested by centrifugation (30 r 8000 rev min^{-1}, 277 K). The use of an enriched media, lo induction temperatures and lower IPTG concentrations served to increase the amount of soluble protein obtaine

Cells were resuspended in ice-cold sonication bu [50 m*M* tris(hydroxymethyl)aminomethane (Tris), 1 phenylmethysulfonyl fluoride (PMSF), 0.1 m*M* dithiothre (DTT)] at a concentration of 1.0 g cells per 1.0 ml bu Lysozyme was added (100 μg ml^{-1}) and the cells incubate room temperature for 30 min. Sonication followed for 6 × intervals at 277 K. Cell debris was pelleted by centrifuga (30 min, 20000 rev min^{-1}, 277 K). Deoxyribonuclease added (20 ng ml^{-1}) and the cell extract filtered (0.02 μm)

The selenomethionine variant of the protein was expres by inhibition of methionine synthesis as described by Dou (1997).

Crude extract was applied at 5 ml min^{-1} to a 7.84 ml an exchange column (Perseptive Biosystems, HQ) equilibrate buffer *A* (50 m*M* Tris pH 7.6, 1 m*M* DTT) and eluted over 500 m*M* NaCl gradient over 15 column volumes. Pro elution was detected on-line at 280 and 420 nm. UDP-ga topyranose mutase was easily identified by its distinct ye colour, a function of the FAD co-factor. A_{420} peaks f subsequent runs were pooled and concentrated u nitrogen pressure in an Amicon ultrafiltration cell. Po fractions from the anion-exchange column were brough 40% $NH_4(SO_4)_2$ saturation (2.4 g ml^{-1}) and applied 5 ml min^{-1} to a 7.84 ml hydrophobic interaction chrom

Figure 1
Alignment displaying the sequence homology between UDP-galactopyranose from *E. coli*, *M. tuberculosis* and *K. pneumonia*. The shaded represents the amino acids common to each enzyme. Overall sequence homology is 40%.

BM14 (λ = 1.001 Å) using a MAR345, 18 cm mode image plate at the European Synchrotron Radiation Facility, Grenoble, France. Data were collected as 145 non-overlapping images in steps of 1° at a crystal-to-detector distance of 242 mm. Diffraction was visible to 2.7 Å, though weak below 2.9 Å (Fig. 3). No significant crystal decay was observed. Data were processed using *DENZO* and *SCALEPACK* (Otwinowski, 1993) yielding 17660 unique reflections. The crystal was indexed in a monoclinic space group with unit-cell dimensions a = 71.12, b = 58.42, c = 97.89 Å, $\alpha = \gamma = 90$, β = 96.38°. Analysis of diffraction data identified systematic absences consistent with space group $P2_1$. Assuming a dimer of molecular mass 43 kDa for the monomer, V_m is 2.35 Å^3 Da^{-1} which falls in the allowed range for protein crystals (Matthews, 1968) and indicates a solvent content of 47%. The data are 92% complete from 15 to 2.9 Å, with an $R_{\rm merge}$ of 5.0% and a redundancy of 2.5. For the highest resolution shell (3.0–2.9 Å), the data have a redundancy of 2.3 with an $R_{\rm merge}$ of 27%.

As no homologous structure is known, we will need to determine phases experimentally. A putative platinum-derivative data set has also been recorded. A crystal of dimensions 0.4 ×0.2 × ~0.03 mm was soaked in 3.5 m*M* K_2PtCl_6 for 9.5 h and back-soaked in cryoprotectant for 2 min before being rapidly transferred to the cold nitrogen stream (100 K). Data were collected as 298 non-overlapping images at 0.5° oscillation on station ID14 EH3 at the ESRF using an MAR CCD with a crystal-to-detector distance of 148 mm. The data are 85% complete to 3.5 Å and have an $R_{\rm merge}$ of 8.1% and a redundancy of 2.8. For the highest resolution shell (3.7–3.5 Å), the data have a redundancy of 2.9 with an $R_{\rm merge}$ of 14%. We were not able to locate the heavy-atom positions with this data. However, an EXAFS experiment did confirm the presence of platinum. In addition, a data set was collected with selenomethionine-variant protein. Data was again collected at 100 K, λ = 0.926 Å, crystal-to-detector distance of 148 mm, using an MAR CCD on ID14 EH3 at the ESRF. A single crystal of dimensions 0.4 × 0.25 × ~0.03 mm was used to record the entire data set in two sweeps of 25 s per 1° oscillation. The data are 94% complete (74% of the anomalous data) to 3.0 Å. $R_{\rm merge}$ is 6.9% and the redundancy is 1.8. For the highest resolution shell (3.16–3.0 Å), the data hav redundancy of 1.9 with an $R_{\rm merge}$ of 13.7%.

We expect structure determination to be aided by presence of two molecules within the asymmetric allowing non-crystallographic averaging. We have exter our studies to the enzymes from *K. pneumonia* and *tuberculosis*. Small crystals of the *K. pneumonia* enzyme been obtained. We expect to solve these structures by m cular replacement based on the *E. coli* structure.

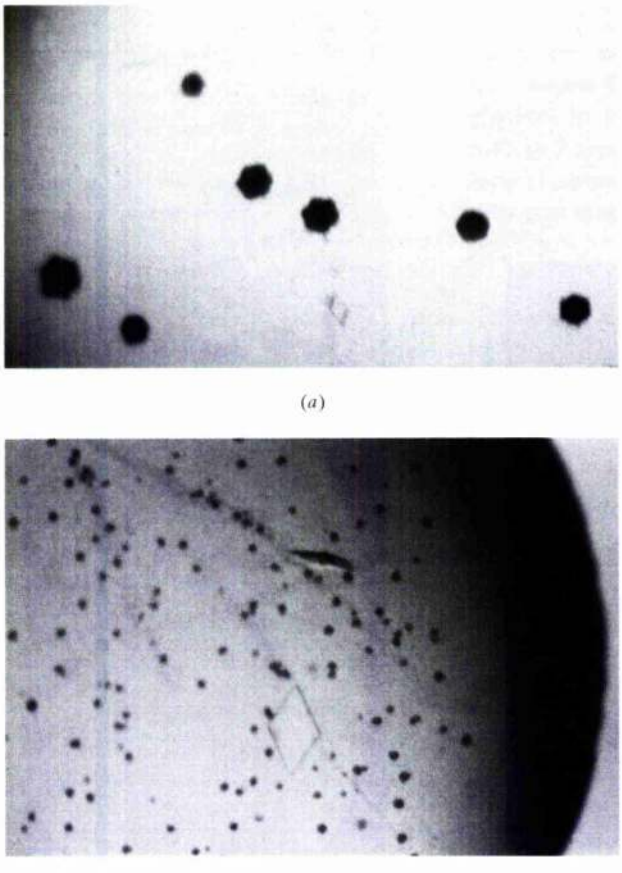
We thank Mike Ferguson, Rob Field and Angela Hunte assistance and helpful discussion throughout this project. thank the Wellcome Trust, the Russell Trust and the R Society for financial support. S. Bird and K. Duncan f Glaxo Wellcome are thanked for the generous donation o mutase clone. We also thank the ESRF for provi synchrotron-radiation facilities. SAM is a University o Andrews St Leonard's College scholar.

References

Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Doublie, S. (1997). *Methods Enzymol.* **276**, 523–530.
Draper, P. (1982). In *The Biology of Mycobacteria*, edited b Ratledge & J. Stanford. London: Academic Press.
Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids Proteins*. Oxford: IRL Press.
Duncan, K. (1997). *Chem. Ind.* **21**, 861–865.
Koplin, R., Brisson, J. & Whitfield, C. (1997). *J. Biol. Chem.* 4122–4128.
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Clo a Laboratory Manual*. New York: Cold Spring Harbor Labora Press.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nassau, P. M., Martin, S. L., Brown, R. E., Weston, A., Monsey McNeil, M. R. & Duncan, K. (1995). *J. Bacteriol.* **178**, 1047–1
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Week* edited by L. Sawyer, N. W. Isaacs & S. Bailey, pp. 5 Warrington: Daresbury Laboratory.
Weston, A., Stern, R. J., Lee, R. L., Nassau, P. M., Monsey, D., Ma S. L., Scherman, M. S., Besra, G. S., Duncan, K. & McNeil, M (1998). *Tubercle Lung Dis.* **78**, 123–131.
Winford, F. G. (1982). In *The Biology of Mycobacteria*, edited b Ratledge & J. Stanford. London: Academic Press.
Young, D. B. (1995). *Odyssey*, **1**, 10–16.

graphy column (Perseptive Biosystems, HP) equilibrated in buffer *B* [50 m*M* Tris, 40% $NH_4(SO_4)_2$, 1 m*M* DTT]. Protein was eluted in a decreasing $NH_4(SO_4)_2$ gradient (100–0%) over 15 column volumes. A_{420} fractions were again combined and concentrated as before. Protein concentration was assayed by the method of Bradford (1976). This purification protocol yields highly homogenous protein for both the native and selenomethionine versions, as judged by sodium dodecyl sulfate and isoelectric focusing polyacrylamide gel electrophoresis. A yield of 8 mg per litre of culture was obtained. Dynamic light-scattering studies showed the protein to be monodisperse with no polydispersity value measurable and a molecular weight of 96 kDa, which is consistent with a dimer of predicted mass 86 kDa. Mass spectroscopic studies (which confirmed selenium incorporation to be greater than 85%) suggest that, as isolated, only 50% of the protein contains both cofactors and about 10% contains neither cofactor.

Crystallization conditions for the native protein were arrived at after screening using the vapour-diffusion hanging-drop methodology (Ducruix & Giegé, 1992). Two crystal forms were obtained. A single crystal of dimensions 0.4 × 0.2 × 0.05 mm was grown over one week from a solution of 10 mg ml^{-1} protein using 10%(*w*/*v*) polyethylene glycol (6K), 100 m*M* HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; pH 7.0) as precipitant. The yellow cr diffracted to 2.7 Å when fresh. The data was indexed *DENZO* (Otwinowski, 1993) in an orthorhombic space g (*P*222, $P2_12_12_1$, $P2_12_12$ or $P222_1$) with cell dimensions *a* *b* = 94, *c* = 134 Å, $\alpha = \beta = \gamma = 90°$. We have not been ab reproduce this form, despite extensive efforts. How crystals grown over 5 d from a solution of 8 mg ml^{-1} pro 100 m*M* HEPES pH 7.6, 20%(*w*/*v*) polyethylene glycol 12% 2-propanol, 0.01 *M* L-cysteine have been consist reproduced. Crystals are yellow diamond-shaped flat plat dimensions 0.04 × 0.02 × ~0.002 mm. The second cr lization conditions are identical for selenomethionine pr with the addition of 1 m*M* DTT. Multiple seeding (2 d for step, 5+ steps being typical) is required to obtain crystals × 0.3 × <0.06 mm (Fig. 2) which are suitable for diffra analysis. Preincubation with both cofactors substan decreases the crystallization time and number of seeding required, although it has little effect on the final crystal Using the in-house Nonius rotating-anode and image- system, diffraction to 3.5 Å is observed. Conditions for protection of the crystals have been developed (15% gly or 15% methane pentanediol) allowing data collecti cryogenic temperatures.



(*a*)

(*b*)

Figure 2
Crystals of UDP-galactopyranose mutase (*a*) before and (*b*) after seeding. The dimensions before seeding are 0.08 × 0.04 × ~0.04 mm and increase to 0.5 × 0.3 × <0.06 mm.

3. Data collection

A single crystal of UDP-galactopyranose mutase dimensions 0.8 × 0.4 × ~0.04 mm was cryoprotected glycerol in mother liquor); soaking took place for app mately 2 min. The crystal was then scooped up with a loop (Hampton Research) and rapidly transferred to the nitrogen stream (110 K). Diffraction data were collecte

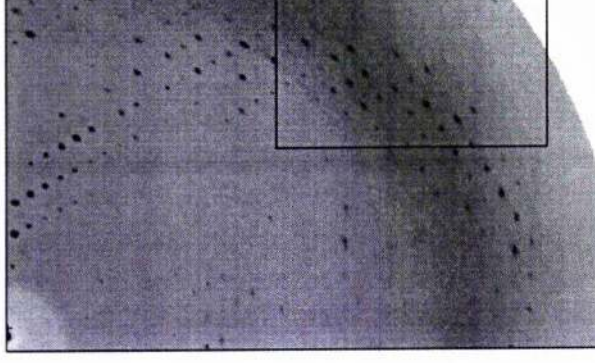


Figure 3
Section of a 1° oscillation diffraction pattern from a crystal of galactopyranose mutase taken at BM14, ESRF, Grenoble recorde an MAR345 image plate. The resolution at the edge of the photo corresponds to 2.8 Å.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary structural studies of dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase (RmlC), the third enzyme o the dTDP-L-rhamnose synthesis pathway, from *Salmonella enterica* serovar Typhimurium

Marie-France Giraud,[a] Fiona M. Gordon,[a] Chris Whitfield,[b] Paul Messner,[c] Stephen A. McMahon[a] and James H. Naismith[a]*

[a]Centre for Biomolecular Sciences, Purdie Building, University of St. Andrews, Fife KY16 9ST, Scotland, [b]Department of Microbiology, University of Guelph, Ontario N1G 2W1, Canada, and [c]Zentrum für Ultrastrukturforschung, Universität für Bodenkultur, A-1180, Vienna, Austria

Correspondence e-mail: naismith@st-andrews.ac.uk

Received 19 October 1998
Accepted 16 November 199

L-Rhamnose is an essential component of the cell wall of many pathogenic bacteria. Its precusor, dTDP-L-rhamnose, is synthesized from α-D-glucose-1-phosphate and dTTP *via* a pathway requiring four distinct enzymes: RmlA, RmlB, RmlC and RmlD. RmlC was overexpressed in *Escherichia coli*. The recombinant protein was purified by a two-step protocol involving anion-exchange and hydrophobic chromatography. Dynamic light-scattering experiments indicated that the recombinant protein is monodisperse. Crystals were obtained using the sitting-drop vapour-diffusion method with ammonium sulfate as precipitant. Diffraction data were collected on a frozen crystal to a resolution of 2.17 Å. The crystal belongs to either space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 71.56$, $c = 183.53$ Å and $\alpha = \beta = 90$, $\gamma = 120°$.

1. Introduction

L-Rhamnose is a key component of the cell wall of many pathogenic bacteria (Shibaev, 1986; McNeil *et al.*, 1990). A full listing of the established primary structures is available in the Complex Carbohydrate Structure Databank (http://www.ccrc.uga.edu). In mycobacteria, L-rhamnose is essential for cell-wall integrity as it connects the inner peptidoglycan layer to the arabinogalactan polysaccharides that are linked to the outer lipid layer of mycolic acids (McNeil *et al.*, 1990). In Gram-negative bacteria, L-rhamnose is often found in the O-antigen part of lipopolysaccharides. This portion of the molecule is often responsible for resistance to complement-mediated serum killing (Joiner, 1988). In *Streptococcus mutans*, the rhamnose-containing polysaccharide has been proposed to be responsible for the colonization of tooth surfaces (Michalek *et al.*, 1984) and adherence to heart, kidney and muscle tissues (Stinson *et al.*, 1980).

L-Rhamnose is incorporated in the mycobacterial cell wall from a nucleoside diphosphate precursor, dTDP-L-rhamnose (Mikusova *et al.*, 1996). In Gram-negative bacteria such as *Salmonella enterica* (Jiang *et al.*, 1991), *Shigella flexneri* (Rajakumar *et al.*, 1994), *Xanthomonas campris* (Köplin *et al.*, 1993) and *Escherichia coli* K12 (Stevenson *et al.*, 1994), as in *Streptococcus mutans* and mycobacteria (Tsukioka *et al.*, 1997; Ma *et al.*, 1997), four enzymes, RmlA, RmlB, RmlC and RmlD, are required to synthesize dTDP-L-rhamnose from α-glucose-1-phosphate and dTTP (Fig. 1).

Because of the importance of L-rhamnose i many pathogenic bacteria, all the enzyme involved in its synthesis are potential target for the design of novel therapeutic inhibitor To this end, we have initiated the structura study of the four enzymes involved in it synthesis. Here, we describe the purificatio crystallization and preliminary structura studies of dTDP-6-deoxy-D-xylo-4-hexulos 3,5-epimerase (RmlC). Bacterial RmlC are nc related in sequence to any other know epimerase.

2. RmlC over-expression and purificatio

The open reading frame of the gene encodin the dTDP-6-deoxy-D-xylo-4-hexulose 3,5 epimerase (RmlC) was amplified by PCR usin primers that incorporated a 5′ *Nde*I site and *Sst*I site to facilitate cloning in the pET30a(+ vector. Expression involves the IPTG-indu cible T7 promoter and ribosome-binding site conferred by the vector, but uses the natura *rmlC* initiation ATG codon. The sequence o the amplified and cloned gene was confirme to be identical to the chromosomal copy. Th expressed protein is therefore strictly identica to the authentic product and carries n extensions or mutations. BL21(DE3) cell transformed with this plasmid were grown a 310 K on Terrific Broth (Maniatis *et al.*, 1982 containing 80 μg ml^{-1} kanamycin until th OD_{600} reached 0.6–0.8. Overexpression wa induced by addition of 1 m*M* IPTG. After 3.5 of culture at 310 K, the cells were harvested b

centrifugation (10 min, 6000*g*, 277 K) and suspended in 100 m*M* NaCl, 2 m*M* DTT, 5 m*M* PMSF, 20 µ*M* lysozyme, 20 m*M* Tris–HCl pH8.

After 30 min of incubation at room temperature, the viscosity of the mixture was decreased by addition of DNAase (20 µg^{-1} ml) and by sonication (four cycles of 30 s interrupted by 1 min periods on ice). After addition of 1 m*M* EDTA, the mixture was centrifuged for 30 min at 20000*g* and 277 K. The supernatant was brought to 20% ammonium sulfate saturation and incubated for 1 h at 277 K. After a second centrifugation (20 min, 20000*g*, 277 K), the supernatant was dialysed against three changes of 2 l of 50 m*M* NaCl, 20 m*M* Tris–HCl pH 8.5. DTT was added to a final concentration of 2 m*M* and the filtered supernatant passed through a POROS-HQ HPLC column (BiocadSprint system). Proteins were eluted with a 50–500 m*M* NaCl gradient. A protein with a molecular weight corresponding to RmlC ($M_r \simeq$ 20.6 kDa) was found in a peak eluted at 250 m*M* NaCl. Fractions corresponding to this peak were pooled, concentrated with an Amicon filter and dialysed against two changes of 1 l of 20 m*M* sodium phosphate pH 7.3. Ammonium sulfate was added gradually to 30% saturation and DTT was added to a final concentration of 2 m*M*. The filtered protein sample was loaded on a POROS high-density phenyl HPLC column (BiocadSprint system) equilibrated in buffer *A* (30% ammonium sulfate, 20 m*M* sodium phosphate pH 7.3). Elution was performed with an increasing gradient of buffer *B* (20 m*M* sodium phosphate pH 7.3); the 20.6 kDa protein was eluted at 75% buffer *B*.

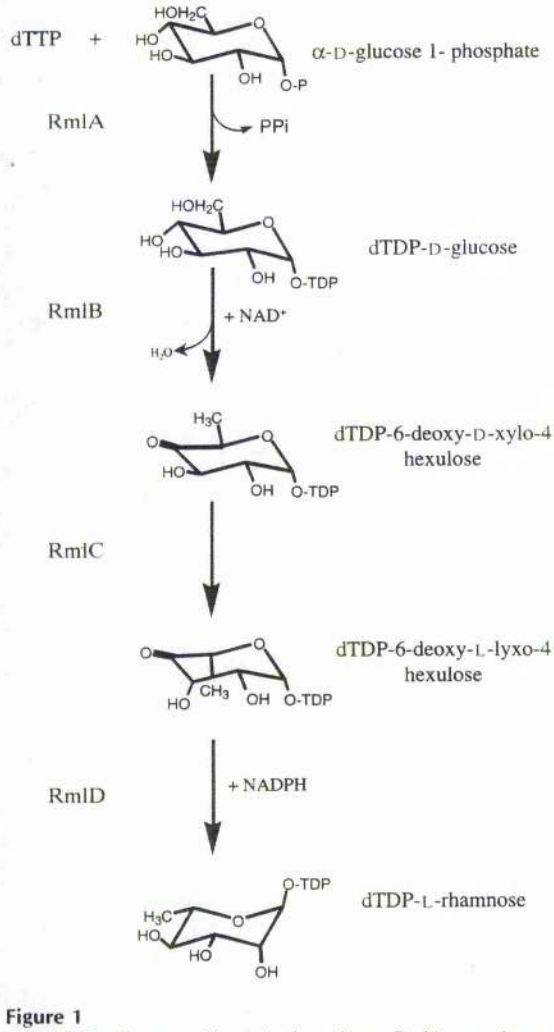


Figure 1
The dTDP-L-rhamnose biosynthetic pathway. RmlA, α-D-glucose-1-phosphate thymidylyltransferase; RmlB, dTDP-glucose 4,6-dehydratase; RmlC, dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase; RmlD, dTDP-6-deoxy-L-xylo-4-hexulose-4-reductase.

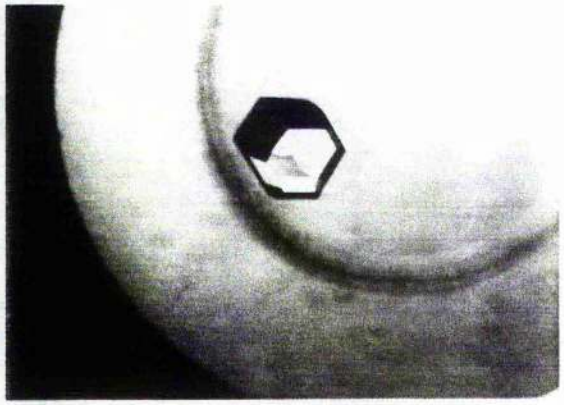


Figure 2
Photograph of an RmlC crystal.

3. Protein analysis

After the two HPLC steps, the protein appeared to be pure as judged on an SDS silver-stained gel. Light-scattering experiments indicated the protein was monodisperse with an apparent molecular weight of 42 kDa, consistent with a dimer. N-terminal sequencing confirmed that the protein was RmlC with an MMIVI N-terminal extremity, with less than 5% of the initiating methionine removed. The final yield of purification was 15 mg l^{-1} Terrific Broth.

Protein concentration was estimated the Bradford method (Bradford, 1976).

4. RmlC crystallization

After the final purification step, RmlC dialysed against three changes of 2 l 25 m*M* Tris–HCl pH 7.75 and concentra to 3.75 mg ml^{-1}, and DTT was then adde a final concentration of 5 m*M*. Crystals w grown with 7 µl protein solution and precipitant (1.6 *M* ammonium sulfate, 0. MES pH 6.1) using the sitting-drop vap diffusion method (Ducruix & Giegé, 19 Hexagonal shaped crystals grew in (Fig. 2).

5. Data collection

Data on frozen crystals (~0.6 × 0.4 0.4 mm) were obtained after a 30 min s in cryoprotectant (20% glycerol, ammonium sulfate, 0.1 *M* MES pH 6 2.65 Å data were collected at 125 K u the Nonius/Macscience DIP2000 im plate. X-rays were generated at a wa length of 1.54 Å from a Nonius FR rotating-anode generator and focused MacScience mirrors. The crystal-to-dete distance was 150 mm. Data were recorde 62 non-overlapping 20 min 1° oscillation

2.17 Å synchrotron data were recorded a frozen crystal at 110 K at the ESRF BM beamline at a wavelength of 0.979 Å usi 30 cm MAR Research image plate. crystal-to-detector distance was 300 Data were recorded as 36 non-overlap 1 min 1° oscillations.

Data on derivatives were recorded Daresbury Laboratory at 125 K usin 30 cm MAR Research image plate beamline 9.5.

The reflections were indexed in a h gonal space group ($a = b = 71$ $c = 183.09$ Å; $\alpha = \beta = 90$, $\gamma = 120°$) *DENZO* and could be scaled in sp groups $P3_121$ or $P3_221$ using *SCALEPA* (Otwinowski, 1993). The V_m va (Matthews, 1968) was 3.3 Å^3 Da^{-1} for molecules per asymmetric unit 2.4 Å^3 Da^{-1} for three molecules per as metric unit. The self-rotation func (Collaborative Computational Pro Number 4, 1994) has no peak for $\kappa =$ but several peaks for $\kappa = 180°$, stro suggesting that the asymmetric unit cont a dimer. Table 1 summarizes the two se data.

Table 1
Data collection.

Values in parentheses refer to the highest resolution shell.

	Data obtained in-house	Data obtained at the ESRF
Resolution (Å)	25–2.65	40–2.17
Highest resolution shell (Å)	2.74–2.65	2.25–2.17
Space group	$P3_121$ or $P3_221$	$P3_121$ or $P3_221$
Unit-cell parameters (Å, °)	$a = b = 71.56$, $c = 183.5$; $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 71.56$, $c = 183.5$; $\alpha = \beta = 90$, $\gamma = 120$
V_m (two molecules per asymmetric unit) (Å^3 Da^{-1})	3.3	3.3
Percentage solvent	64.4	64.4
Unique reflections	16155	27346
I/σ	15.1 (2.3)	19.5 (1.9)
Average redundancy	3.0 (2.0)	2.2 (2.09)
Data completeness (%)	97.1 (86.2)	92.1 (80.7)
R_{merge}† (%)	7.5 (27.3)	4.7 (38.4)

† $R_{merge} = \sum\sum |I(h)_j - \langle I(h)\rangle| / \sum\sum I(h)_j$ where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

6. Derivatization

A partial data set (50%) was recorded at Daresbury Laboratory on a derivative obtained with a 1 h soak in 1 m*M* $HgCl_2$. This derivative had four Hg atoms bound per unit cell with a phasing power of 1.6 and Cullis *R* factors of 0.64 for acentric data and 0.5 for centric data. We have produced other mercury derivatives and selenomethionine-enriched protein and will determine the RmlC structure shortly.

The use of the CCLRC Daresbury Laboratory UK and ESRF-Grenoble facilities is gratefully acknowledged. We thank The Wellcome Trust for an international travel grant.

References

Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins, a Practical Approach*. Oxford: IRL Press.
Jiang, X.-M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991). *Mol. Microbiol.* **5**, 695–713.
Joiner, K. A. (1988). *Ann. Rev. Microbiol.* **42**, 201–230.
Köplin, R., Wang, G., Hötte, B., Priefer, U. B. & Pühler, A. (1993). *J. Bacteriol.* **175**, 7786–7792.
Ma, Y., Mills, J. A., Belisle, J. T., Vissa, V., Howell, M., Bowlin, K., Sherman, M. S. & McNeil, M. (1997). *Microbiology*, **143**, 937–945.
McNeil, M., Daffé, M. & Brennan, P. J. (1990). *J. Biol. Chem.* **265**, 18200–18206.
Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Michalek, S. M., Morisaki, I., Gregory, R. L., Kimura, S., Harmon, C. C., Hamada, S., Kotani, S. & McGene, J. R. (1984). *Protides Biol. Fluids Proc. Colloq.* **32**, 47–52.
Mikusova, K., Mikus, M., Besra, G., Hancock, I. & Brennan, P. J. (1996). *J. Biol. Chem.* **271**, 7820–7828.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend*, edited by L. Sawyer, N. Isaacs & S. W. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Rajakumar, K., Jost, B. H., Sasakawa, C., Okada, N., Yoshikawa, M. & Adler, B. (1994). *J. Bacteriol.* **176**, 2362–2373.
Shibaev, V. N. (1986). *Adv. Carbohydr. Chem. Biochem.* **44**, 277–339.
Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M. Redmond, J. W., Lindquist, L. & Reeves, P. (1994) *J. Bacteriol.* **176**, 4144–4156.
Stinson, M. W., Nisengard, R. J. & Bergey, E. J. (1980). *Infect. Immun.* **27**, 604–613.
Tsukioka, Y., Yamashita, Y., Nakano, Y., Oho, T. & Koga, T. (1997) *J. Bacteriol.* **179**, 4411–4414.

Acta Cryst. (1998). D**54**, 1023–1025

Crystallization of succinylated concanavalin A bound to a synthetic bivalent ligand and preliminary structural analysis

DAVINA N. MOOTHOO,[a] STEPHEN A. MCMAHON,[a] SARAH M. DIMICK,[b] ERIC J. TOONE[b] AND JAMES H. NAISMITH[a]* *at* [a]*Centre for Biomolecular Science, Purdie Building, The University, St Andrews KY16 9ST, Scotland, and* [b]*Departments of Biochemistry and Chemistry, Duke University, North Carolina, NC 27708–0346, USA. E-mail: naismith@st-and.ac.uk*

(*Received 2 February 1998; accepted 9 March 1998*)

Abstract

Crystals have been obtained of succinylated concanavalin A complexed to a novel bidentate synthetic ligand. The crystals are the first example of a lectin with a synthetic multivalent ligand and the first report of crystallization of succinylated concanavalin A. The crystals were obtained by sitting-drop vapour diffusion equilibrating with a solution of 20% polyethylene glycol, pH 5, 293.5 K. Crystals are orthorhombic, belonging to space group $C222_1$ with unit-cell dimensions of $a = 99.1$, $b = 127.4$, $c = 118.9$ Å. The asymmetric unit contains a dimer, with over 65% of the volume occupied by water. The ligand cross links concanavalin A monomers. Succinylated concanavalin A is known to be a dimer in solution, yet it is found as the typical concanavalin A tetramer in the crystal. The contacts holding together the tetramer appear extensive and suggest that a fine balance between dimer and tetramers exists. Data to 2.65 Å have been collected and the structure determined by the molecular replacement method.

1. Introduction

Lectins comprise a varied family of sugar-binding proteins; they are found in all types of organisms. Plant lectins in particular have been of intense interest, because they exhibit exquisite specificity for oligosacccharides and unlike their mammalian counterparts are much more tractable to characterization by a broad range of biophysical techniques. Thus, these proteins, such as the lectin from *Canavalia ensiformis*, concanavalin A (con A), have served as models for more complex species (Rini, 1995). In contrast to the exquisite selectivity at the oligosaccharide level, the proteins are relatively promiscuous at the monosaccharide level, falling into two broad classes; mannose specific and galactose specific. The affinity of the lectin for monosaccharides is correspondingly lower, of the order $K_a = 1 \times 10^3\ M^{-1}$ compared with $K_a = 1 \times 10^6\ M^{-1}$ for oligosaccharides (Chervenak & Toone, 1995; Mandal *et al.*, 1994; Toone, 1994). The advent of calorimetric data on lectin (especially con A) carbohydrate complexes and the realisation that *ab initio* modelling methods were failing to accurately model protein–carbohydrate interactions has reinvigorated the structural study of lectin–carbohydrate complexes. A number of oligosaccharide–protein complexes have been determined (Rini, 1995), most recently the structures of con A with its cognate trisaccharide and pentasaccharide (Naismith & Field, 1996; Moothoo & Naismith, 1998).

Protein–saccharide interactions are appealing therapeutic targets in many diseases, particularly those involving infection and inappropriate immune response (Dwek, 1996). Therapeutics designed for this task would principally be aimed at disrupting or interfering with molecular recognition processes rather than being cytotoxic. Oligosaccharides are poor therapeutics, however, as they are too polar for satisfactory uptake. Monosaccharides, however, bind too weakly to be of value. One possible solution to this conundrum has been the use of polyvalent ligands. This is based on the observation that in nature many carbohydrate proteins appear to function as oligomers. Chemical synthesis has produced families of such multivalent ligands, which on the basis of agglutination assays appear to have dramatic results (Kanai *et al.*, 1997; Sigal *et al.*, 1996). Several structures are now known of lectins complexed to naturally occurring carbohydrates that cross link the protein in the crystalline phase (Bourne *et al.*, 1994; Dessen *et al.*, 1995; Wright, 1992; Wright & Hester, 1996). Important questions remain as to the thermodynamic basis of these polyvalent interactions (Roy, 1996; Toone, 1994). As part of a program to combine calorimetery, crystallography and chemical synthesis we report the crystallization and structure determination of the first synthetic multivalent ligand–lectin complex.

2. Crystallization and X-ray data collection

Succinylated con A was purchased from Sigma (Poole, UK). The synthesis of the bidentate ligand [1,3-di-(*N*-propyloxy-α-D-mannopyranosyl)-carbomyl 5-methylazido-benzene, Fig. 1] will be described elsewhere. Crystallization trials were performed by the means of the sitting-drop vapour-diffusion method (Ducruix & Giegé, 1992). A solution of succinylated con A (1.2 m*M*)/ligand (18 m*M*) was prepared in 20 m*M* Tris pH 7, 100 m*M* NaCl, 1 m*M* $CaCl_2$ and 1 m*M* $MnCl_2$ and equilibrated against 20% polyethylene glycol (M_r = 6000), 100 m*M* citric acid, pH 5 in sitting-drop trays (Charles Supper, USA) at 293.5 K. Small crystals of dimensions 0.1 × 0.2 × 0.2 mm were initially obtained. Optimization of the ligand concentration (5 m*M*) yielded block-shaped crystals of up to 1.0 × 0.8 × 0.6 mm. Crystal growth was complete in 14 d. 12 crystals were examined, mounted in a thin-walled glass capil-

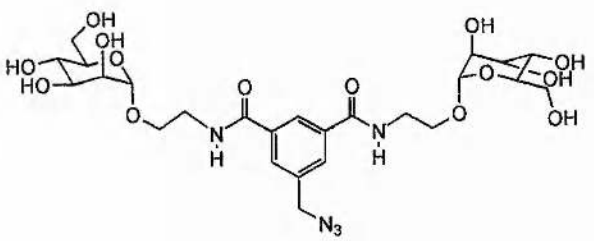


Fig. 1. The bivalent ligand [1,3-di-(*N*-propyloxy-α-D-mannopyranosyl)-carbomyl 5-methylazido-benzene] which cross links con A.

ISSN 0907-4449

lary, by exposure to X-rays. The resolution limit varied from 2.65 to 3.5 Å. All diffraction data were collected at 293.5 K from a single crystal (0.6 × 0.5 × 0.3 mm) which diffracted to the highest resolution. Larger crystals did not give higher resolution data. All data were recorded on the Nonius/MacScience DIP2000 dual image plate. X-rays were generated using a Enraf–Nonius FR951 rotating-anode generator and focused using the MacScience mirror system, through a 0.5 mm collimator. Data were collected as 92 non-overlapping 25 min 1 oscillations with a crystal-to-detector distance of 140 mm. The resolution limit of the data was 2.65 Å and no significant crystal decay was observed during data collection. The programs *DENZO* and *SCALEPACK* (Otwinowski, 1993) were used to process the data. The crystal was indexed in a centred orthorhombic space group with unit-cell dimensions $a = 99.1$, $b = 127.4$, $c = 118.9$ Å. Analysis of diffraction data identified systematic absences consistent with space group $C222_1$. A dimer of molecular mass 49 kDa, gives rise to a V_m (Matthews, 1968) of 3.8 Å^3 Da^{-1} and indicates a solvent content of over 65%. The data are 96% complete from 26 to 2.65 Å, with an R_{merge} 7.2% with a redundancy of 2.2 and a total of 84% of possible data are greater than 1σ. For the high-resolution shell (2.75–2.65 Å), the corresponding values are 98% complete, R_{merge} 20.2%, redundancy 2.2 and 83% greater than 1σ.

3. Structure solution

The structure was solved by molecular replacement, using *AMoRe* (Navaza, 1994) as implemented in the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). The conventional con A dimer (monomers *A* and *B*) from the trimannoside complex of con A (1CVN) stripped of metals, waters and sugars was used as the search model. The rotation function found a single solution with a correlation coefficient of 0.28 and the translation function produced a final solution with a correlation coefficient of 0.84. A translation search in *C*222 gave no solution. In order to visualize the interaction of the ligand with each protein monomer, the asymmetric unit was redefined such that one monomer of the con A dimer (monomer *A*) was linked *via* the ligand to another monomer (monomer *C*). Monomer *C* was generated by applying the symmetry transformation $(\frac{1}{2} - x, \frac{1}{2} - y, z + \frac{1}{2})$ to monomer *B* (Fig. 2). Other crystallographic operators generate the 'typical' con A tetramer. The initial free *R* factor was 38% to 2.65 Å. Rigid-body refinement of the monomers gave a free *R* factor of 30%. Clear density is visible for the metal ions and ligand. The ligand is currently being built into this electron density and a stereochemical dictionary constructed. Once the ligand has been built into density, refinement will recommence. A fully refined structure will be reported elsewhere.

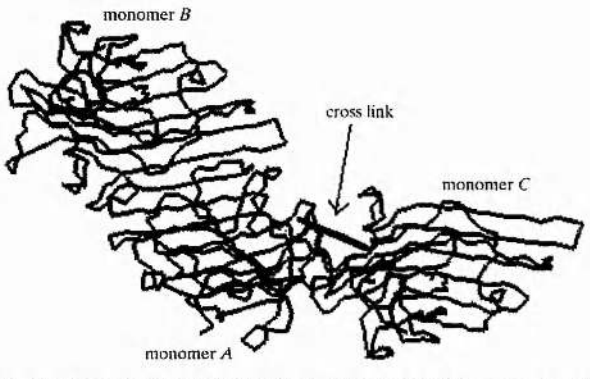


Fig. 2. The conventional dimer of con A is formed by monomer *A* and monomer *B*, the ligand cross links monomer *A* to monomer *C*. Monomer *C* is related by the crystallographic transformation $(\frac{1}{2} - x, \frac{1}{2} - y, z + \frac{1}{2})$ to monomer *B*. The cross-linking ligand is shown a straight line between the two sugar-binding sites.

That succinylated con A is found as a tetramer in this crystal is puzzling; the electron density shows evidence of succinylation on two lysine residues. Dimerization of the dimers occludes 4300 Å^2 of surface area which compares to 5200 Å^2 for native con A (Deacon *et al.*, 1997; Reeke *et al.*, 1975) and 4800 Å^2 for the mannose con A complex (Naismith *et al.*, 1994). There are 120 protein–protein contacts less than 4.0 Å between the two dimers in this structure, compared with 260 for the native and 160 for the mannose complex. Although this tetramer is clearly less tightly packed than the other two, by a standard set of standard crystallographers' criteria (buried surface area and contacts) one would identify the protein to be a functional tetramer. However, solution studies show unambiguously that succinylated con A is a dimer (Gunther *et al.*, 1973). It appears that the energy barrier between dimers and tetramers for succinylated con A is small enough that crystal-packing forces can drive formation of the tetramer. We feel this result poses an interesting test for those who wish to devise automatic methods of identifying protein oligomerization state from static crystal structures.

JHN acknowledges BBSRC support (B08307) and we thank John Helliwell for encouragement.

References

Bourne, Y., Bolgiano, B., Liao, D. L., Strecker, G., Cantau, P., Herzberg, O., Feizi, T. & Cambillau, C. (1994). *Nature Struct. Biol.* **1**, 863–870.
Chervenak, M. C. & Toone, E. J. (1995). *Biochemistry*, **34**, 5685–5695.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Deacon, A., Gleichmann, T., Kalb (Gilboa), A. J., Price, H., Rafferty, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **53**, 4305–4312.
Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F. & Sacchettini, J. C. (1995). *Biochemistry*, **34**, 4933–4942.
Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins*. Oxford: IRL Press.
Dwek, R. A. (1996). *Chem. Rev.* **96**, 683–720.
Gunther, G. R., Wang, J. L., Yahara, I., Cunningham, B. A. & Edelman, G. (1973). *Proc. Natl Acad. Sci. USA*, **70**, 1012–1016.
Kanai, M., Mortell, K. H. & Kiessling, L. L. (1997). *J. Am. Chem. Soc.* **119**, 9931–9932.
Mandal, D. K., Kishore, N. & Brewer, C. F. (1994). *Biochemistry*, **33**, 1149–1156.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Moothoo, D. N. & Naismith, J. H. (1998). *Glycobiology*, **8**, 173–181.
Naismith, J. H., Emmerich, C., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Raftery, J., Kalb, A. J. & Yariv, J. (1994). *Acta Cryst.* D**50**, 847–858.
Naismith, J. H. & Field, R. A. (1996). *J. Biol. Chem.* **271**, 972–976.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. (1993). In *Proceedings of the CCP4 study weekend: Data collection and processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Reeke, G. N. J., Becker, J. W. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1525–1547.
Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.
Roy, R. (1996). *Curr. Opin. Struct. Biol.* **6**, 692–702.
Sigal, G. B., Mammen, M., Dahmann, G. & Whitesides, G. M. (1996). *J. Am. Chem. Soc.* **118**, 3789–3800.
Toone, E. J. (1994). *Curr. Opin. Struct. Biol.* **4**, 719–728.
Wright, C. S. (1992). *J. Biol. Chem.* **267**, 14345–14352.
Wright, C. S. & Hester, G. (1996). *Structure*, **4**, 1339–1352.